Working Paper

Department of Finance

Economic and Fiscal Policy Branch

Working papers are circulated in the language of preparation, to make analytic work undertaken by the staff available to a wider readership. They have received only limited evaluation and do not reflect the views of the Department of Finance. Comments on the working papers are invited and may be sent to the author(s).

A Guide to Applied Modern Macroeconometrics

by
Alain Paquet*

Working Paper No. 94-05

Fiscal Policy and Economic Analysis Branch
Department of Finance, Canada

*Department of Economics, and Research Center on Employment and Economic Fluctuations, Université du Québec à Montréal, P.O. Box 8888, Station Centre-ville, Montréal, Québec H3C 3P8. This is a revised version of an earlier document prepared for a course given at the Department of Finance, Canada, in the Spring 1994.

Working papers are intended to make analytic work undertaken in the Department of Finance available to a wider readership. They have received only limited evaluation and do not necessarily reflect the views of the Department of Finance. Comments on the working papers are invited and may be sent to the author(s). They are circulated in the language of preparation.

Table of Content

1. INTRODUCTION. 1-1

2. TOOLS FOR MODEL SELECTION IN APPLIED ECONOMETRICS. 2-1

2.1 Introduction. *2-1*

2.2 The General Linear Model, its Conventions, and Some Commonly Used Estimators. *2-2*

2.2.1 Review of the general linear model, its conventions, and the OLS estimator. *2-2*

2.2.2 Review of the problems occurring from the violations of the underlying conventions and alternative estimators. *2-3*

2.2.2.1 Violation of conventions 1 and 2 *2-4*

2.2.2.1.1 Misspecification and simultaneity bias 2-4

2.2.2.1.2 Instrumental variable estimators 2-6

2.2.2.2 Violation of convention 3 *2-7*

2.2.2.2.1 Nonspherical error term 2-7

2.2.2.2.2 Feasible generalized estimator and heteroskedasticity and serial correlation consistent estimator of the variance-covariance matrix of the regression coefficients. 2-8

2.2.2.3 Violation of convention 4: non-normal error term. *2-10*

2.3 Statistical Inadequacy and Specification Tests in Econometric Models. *2-11*

2.3.1 Assessment of an empirical model by variable addition: a general principle. *2-11*

2.3.1.1 Augmenting the conditional mean *2-11*

2.3.1.1.1 The RESET test 2-14

2.3.1.1.2 The Chow test for constancy of the coefficients 2-15

2.3.1.1.3 The predictive failure test 2-16

2.3.1.1.4 The serial correlation test 2-16

2.3.1.1.5 The Hausman test of exogeneity 2-17

2.3.1.1.6 Plosser, Schwert and White differencing test 2-18

2.3.1.2 Augmenting the conditional variance *2-19*

2.3.1.3 Augmenting the higher conditional moments: a test of normality *2-20*

2.3.1.4 Joint tests *2-22*

2.3.2 Statistical inadequacy tests for structural equations *2-24*

2.3.3 Non-nested tests and encompassing tests *2-27*

2.3.4 Information criteria for selecting regressors *2-29*

2.4 An Econometric Approach to Model Selection: Hendry's Methodology. 2-31

2.5 Summary. 2-34

3. THE CHARACTERIZATION OF UNIVARIATE TIME-SERIES. 3-1

3.1 Introduction. 3-1

3.2 Identifying, Estimating and Diagnostic Checking. 3-1

3.2.1 Introduction 3-1

3.2.2 Definitions and theoretical properties of stationary stochastic processes 3-2

3.2.3 Definitions and theoretical properties of some nonstationary stochastic processes 3-12

3.2.4 Practical tools for the identification of stochastic processes 3-13

3.2.4.1 Sample autocorrelations and sample partial autocorrelations 3-14

3.2.4.2 Recursive sample autocorrelations 3-16

3.2.4.3 Seasonality and other regular patterns in sample autocorrelations and sample partial autocorrelations 3-16

3.2.4.4 Hannan and Risssanen's methodology for identification 3-18

3.2.5 Estimation of a stationary ARMA process 3-19

3.2.6 Diagnostic checking on the residuals of an estimated ARMA process 3-20

3.2.6.1 ACF and PACF of the residuals 3-20

3.2.6.2 Portmanteau statistics 3-20

3.2.6.3 LM tests of extra coefficients 3-21

3.2.6.4 Other specification tests 3-22

3.2.7 Example: the Canadian exchange rate 3-23

3.2.7.1 Identification of the stochastic process: ACF , PACF and other visual tools 3-23

3.2.7.2 Identification of its stochastic process. Information criteria analysis 3-26

3.2.7.3 Estimation of a stochastic process 3-27

3.2.7.4 Diagnostic checking: ACF and PACF of the residuals 3-28

3.2.7.5 Diagnostic checking and test statistics 3-29

3.3 The Issue of Nonstationarity and Unit Roots in Macroeconomic Time Series 3-31

3.3.1 Introduction 3-31

3.3.2 Deterministic versus stochastic trend 3-31

3.3.3 Unit root tests 3-34

3.3.3.1 Cochrane's variance ratio test 3-34

3.3.3.2 Simple and augmented Dickey-Fuller tests 3-35

ABSTRACT

The 1980s and early 1990s have set applied macroeconometrics on a new course because of advances in econometric theory and the development of ever more powerful computers and software. The object of this guide is to provide an operational survey of some of the most useful tools encountered in modern applied macroeconomic research by making accessible a step-by-step presentation of many of these techniques. With an emphasis clearly practical in scope, it seeks to describe the motivation and the role of the techniques, to explain the steps one must perform to use them properly, and to point out the caveats one must be aware of when interpreting the results obtained from their application. Formal proofs are avoided as a rule, but heuristic and intuitive explanations of the statistical and econometric theoretical concepts underlying these methods are given. This is important to understand the strengths and weaknesses of these methods to do good applied work. Practical examples are also included to illustrate the use of the techniques, and various tables of critical values are assembled in the chapter's appendices.

Chapter 2 first intuitively reviews the foundation of basic econometrics and highlights the role of the underlying statistical conventions that justify the use of least squares estimation methods. It then presents various tests that serve to look for evidence of violations of these conventions with respect to misspecifications, serial correlation, heteroskedasticity, nonnormality, and instability. Both individual and joint tests are discussed in the context of individual equations. Extensions to structural equations and non-nested tests are also presented. Finally the use of these tests is motivated within Hendry's general-to-specific approach.

The third chapter addresses the issues of identification, estimation and diagnostic checking of univariate ARIMA time-series models. Both visual and formal statistical tools are presented. Topics relating to the nonstationarity or integration of time series are covered. These include several unit root tests, their statistical properties, things that need to be known to make proper use of them, and the meaning of unit roots in macroeconomics.

Chapter 4 builds on the previous chapters to address the construction of dynamic econometric models both for single equations and equations systems (including VAR and SVAR models). Various equivalent reparameterisations of dynamic models are discussed. Particular consideration is given to the issues of spurious versus fundamental long-run (or cointegrating) relationships between integrated variables. The consequences of building empirical models with integrated variables for identification, estimation and inference are discussed. Subjects treated in this chapter include: residual based tests of non cointegration; optimal and efficient estimation of cointegrating relationships in single equations and in systems; Johansen's methods for testing the dimension of a cointegrating space and for estimation; and Granger non-causality tests.

Chapter 5 concludes by discussing how the uses of modern macroeconometric tools fit together within a self-contained and systematic econometric approach.

RÉSUMÉ

Les progrès en théorie économétrique et le développement d'ordinateurs et de logiciels plus puissants dans les années 80 et le début des années 90 ont permis à la macroéconométrie appliquée de s'engager sur de nouveaux chemins. Ce guide vise à fournir un survol opérationnel de certains outils utiles pour la recherche appliquée de pointe en macroéconomie en rendant accessible une présentation par étape de plusieurs de ces techniques. Son emphase est clairement pratique puisqu'il cherche à décrire la motivation et le rôle des techniques, à expliquer les étapes nécessaires pour les appliquer correctement et à faire ressortir les *caveats* reliés à leur application. Des preuves rigoureuses sont évitées, mais des explications heuristiques et intuitives des concepts statistiques et économétriques théoriques sous-jacents sont présentées. Il est important de développer cette intuition afin de comprendre les forces et les faiblesses des méthodes pour faire du bon travail empirique. Des applications pratiques sont aussi incluses pour illustrer l'utilisation des techniques et de nombreuses tables de valeurs critiques sont rassemblées en appendice des chapitres.

Le chapitre 2 revoit intuitivement les fondements de l'économétrie de base et fait ressortir les conventions statistiques sous-jacentes qui justifient l'usage des méthodes d'estimation des moindres carrés. Il présente ensuite plusieurs tests qui cherchent à établir des éléments d'évidence empirique révélant des violations des conventions en regard de problèmes de spécification, de dépendance temporelle, d'hétéroscédasticité, de non normalité et d'instabilité. Des tests d'hypothèses individuelles et jointes sont discutés pour des équations individuelles. Des extensions aux cas d'équations structurelles et des tests d'hypothèses non-emboîtées sont aussi présentées. L'utilisation de ces tests est finalement motivée dans le cadre de l'approche de Hendry qui recommande ces tests pour passer d'un modèle empirique général à un modèle plus spécifique et statisquement adéquat.

Le troisième chapitre s'intéresse aux questions d'identification, d'estimation et de tests de diagnostics pour des modèles univariés de séries chronologiques de type ARIMA. Des outils visuels et statistiques formels sont présentés. Les sujets pertinents aux séries chronologiques non-stationnaires, ou intégrées sont discutés. Ceux-ci incluent: les tests de racines unitaires, leurs propriétés statistiques, les éléments nécessaires pour les appliquer convenablement et la signification des racines unitaires en macroéconomie.

Le chapitre 4 utilise les développements des chapitres précédents pour discuter l'élaboration de modèles économétriques dynamiques dans le cadre d'équations individuelles et de systèmes d'équations (incluant les modèles VAR et SVAR). Plusieurs reparamétrisations équivalentes de modèles dynamiques sont présentées. Une attention particulière est accordée aux questions touchant les relations empiriques fictives versus les relations fondamentales de long terme (ou de cointégration) entre des variables intégrées. Les conséquences pour l'identification, l'estimation et l'inférence occasionnées par la présence de variables non-stationnaires dans les modèles empiriques sont abordées. Les sujets traités incluent; les tests de non-cointégration basés sur des résidus, l'estimation efficiente et optimale de relations de cointégration dans des équations individuelles et dans des systèmes d'équations, la méthode de Johansen pour tester la dimension de l'espace de cointégration et l'estimation, et les tests de non causalité de Granger.

Finalement, le chapitre 5 conclut par une présentation schématique d'une méthodologie économétrique globale et systématique qui fait usage des outils macroéconométriques modernes.

3.3.3.3 Phillips and Perron tests *3-45*

3.3.3.4 Statistical sizes of the tests and nature of the underlying stochastic process *3-46*

3.3.3.4.1 A comparison of ADF tests and Phillips-Perron tests 3-46

3.3.3.4.2 Unit root tests and negative MA component in the innovations 3-46

3.3.3.5 Power of unit root tests *3-48*

3.3.3.6 Unit root tests and structural breaks *3-48*

3.3.3.7 Unit root tests and officially seasonally adjusted series *3-49*

3.3.4 Strategy of unit root testing and applications *3-50*

3.3.5 Cochrane variance ratios and unit root tests for the log Canadian exchange rate and its first difference *3-53*

3.3.6 Unit roots in macroeconometrics *3-55*

3.4 Summary. ***3-57***

3.5 Appendix Tables. ***3-59***

4. DYNAMIC ADJUSTMENT MODELS, LONG-RUN RELATIONSHIPS AND COINTEGRATION. **4-1**

4.1 Introduction. ***4-1***

4.2 Dynamic Models And Error-Correction Models. ***4-1***

4.3 Relationships Between Intgrated Series in Individual Equations: Spurious Regression vs. Cointegration. ***4-10***

4.3.1 Spurious regressions in econometrics *4-10*

4.3.2 Cointegrating relationships *4-11*

4.3.3 Regressions and cointegration in individual equations *4-13*

4.3.3.1 The case of two variables. *4-13*

4.3.3.1.1 Estimation in static regressions 4-13

4.3.3.1.2 Inference about cointegration in static regressions 4-15

4.3.3.1.3 Cointegration and dynamic models 4-18

4.3.3.1.4 Optimal estimators of cointegrating coefficients with asymptotic standard distribution in individual equations with endogenous regressors 4-21

4.3.3.2 Generalization to the N variables case. *4-23*

4.3.3.3 Inference about parameters of integrated regressors. *4-24*

4.3.3.4 How to deal with I(1) regressors in regressions: the case of the permanent income theory of consumption *4-28*

4.3.3.4.1 Application 1 4-28

4.3.3.4.2 Application 2 4-29

4.3.3.4.3 Application 3 4-29

4.3.3.4.4 Application 4 4-30

4.3.3.4.5 Application 5 4-30

4.3.3.4.6 Application 6 4-31

4.3.4 Example: inflationary expectations and rationality I 4-32

4.4 Multivariate Time Series Processes. **4-37**

4.4.1 Introduction 4-37

4.4.2 Vector autoregressive models (VAR) 4-37

4.4.2.1 The case of stationary regressors. 4-37

4.4.2.1.1 Set-up, identification, and choice of the order of a VAR 4-37

4.4.2.1.2 Estimation and diagnostic checking 4-39

4.4.2.1.3 Applied tools specific to VAR models 4-40

4.4.2.1.4 Issues in structural VAR's 4-45

4.4.2.2 Generalization to the case of I(1) variables. 4-49

4.4.3 Vector error-correction models (VECM) and cointegration in dynamic equation systems 4-52

4.4.3.1 The set-up and its usefulness 4-52

4.4.3.2 Johansen's estimation and testing procedure. 4-58

4.4.3.2.1 FIML estimation of a VECM 4-58

4.4.3.2.2 Determination of the cointegrating rank 4-60

4.4.3.2.3 Hypothesis testing about the presence or absence of a linear trend 4-64

4.4.3.2.4 Hypothesis testing about the adjustment matrix α 4-65

4.4.3.2.5 Hypothesis testing about the basis of the cointegrating space β 4-67

4.4.3.2.6 Granger non-causality tests in levels VAR and in ECM 4-69

4.4.3.2.7 Johansen's procedure and other optimal estimators of cointegrating relationships 4-70

4.4.3.3 Example:inflationary expectations and rationality II 4-71

4.5 Summary. **4-77**

4.6 Appendix Tables. **4-79**

5. CONCLUSIONS. **5-1**

REFERENCES **R-1**

1. INTRODUCTION.

Twenty-five years ago, a good understanding of econometrics required reasonable knowledge of linear algebra, of multivariate statistics, and of basic methods in time-series analysis. While this is still true today, further specialisation has become the thrust of econometric methodology. Micro-econometricians need a solid knowledge of maximum likelihood estimation and of non-parametric estimation techniques, as well as the skills to treat very large data banks. Macro-econometricians need to deal with statistical issues implied by nonstationary regressors, the development of dynamic models, and the generalized method of moments.

With the advent of more powerful computers and the developments of more sophisticated mathematics, econometric methods have certainly become more involved and may seem to be almost too complicated to tackle proficiently by non-specialists. At the same time, user-friendly and non-user-friendly econometric softwares tend to offer ever more options and tests that can provide many statistics, but without providing explanations about the meaning and the relevance of these numbers. While econometrics may well be an art, it is possible to draw a coherent methodology for selecting and testing models that can make sense to non-specialists.

In the 1980's, topics such as model selection, diagnostic tests, stationarity, cointegration and error correction models have received much attention in theoretical econometric research. The use of many of these techniques is now becoming the norm for empirical work. Furthermore, these state-of-the-art techniques must not be seen just as sophisticated tools that we employ on limited quality data to get the same results that we would have obtained with the unquestioned and old-fashioned application of OLS estimation and the usual F- or t-tests of significance. On the contrary, they are important to reveal a better understanding and characterisation of the data, to build better forecasting equations, and to discriminate eventually between competing theories. It is therefore important to develop an intuitive and usable understanding of the new methods.

The object of this reference manual is to provide an operational survey of some of the most useful tools encountered in modern applied macroeconomic research. Its emphasis is clearly practical in scope. It seeks to describe the motivation and the role of the techniques, to explain the steps one must perform to use them properly, and to point out the *caveats* one must be aware of when interpreting the results obtained from their application. One aim is to understand the strengths and weaknesses of these methods to do good applied work. Heuristic and intuitive explanations of the statistical and econometric theoretical concepts underlying these methods are often given. More technical demonstrations are avoided as a rule, while references to other material are included for the interested readers.[1] Throughout the text, examples are also occasionally provided to illustrate the use of these techniques.

Many available computer softwares for statistics and econometrics perform many, if not most of the tools discussed in this guide.[2] In particular, the newest version of MicroTSP for Windows, called Econometric Views 1.0, appears well suited to do modern macroeconometric analysis. Several built-in commands allow the direct application of techniques. This software comes also with a fairly powerful command language that allows one to program his or her favorite procedures. Vincent Lépine (from the Department of Finance) and I have begun to write a library of programs for EViews in the spirit of this guide. These programs are available upon request.

Compared to a first version that was used in a series of lectures given at the Department of Finance in the Spring 1994, this guide has undergone fairly significant revisions, especially in chapters 3 and 4, while I was visiting the Federal Reserve Bank of Minneapolis. I would like to acknowledge Benoît Robidoux's encouragement and very helpful suggestions. I am thankful for his comments and suggestions, as well as those made by Christian Dea, Vincent Lépine, Richard Stuart, Bing-Sun Wong (all from the Department of Finance), and by Liliane Brouillette (MA graduate in mathematics-statistics and in economics, UQAM), Doug Hodgson (Ph.D. student in economics, Yale University), and Jean Laferrière (MA graduate in economics, UQAM). Other thanks go to the participants of the Department of Finance's lectures and to the many students in the macroeconometrics course that I have thought at UQAM. The usual disclaimer applies.

A.P.

---

[1] As I was completing the current revisions, a new comprehensive graduate textbook by Hamilton (1994) had just come out. While specific references have not yet generally been incorporated in this version of the guide, I recommed this book as an excellent reference for other topics and for proofs and more formal expositions of much of the material covered here.

[2] A non-exhaustive list of these softwares includes: GAUSS, Microfit, MicroTSP, PC-GIVE, RATS, SAS, and TSP International. These are all available in PC versions. Many have also versions for workstation, and some have mainframe versions.

2. TOOLS FOR MODEL SELECTION IN APPLIED ECONOMETRICS.

> "Until the model characterizes the data generation process, it seems rather pointless trying to test hypotheses of interest in economic theory."
>
> David Hendry (1979)

2.1 Introduction.

In applied research, it is very important to submit one's empirical model to a battery of specification tests. In many papers, estimation techniques that are employed, assume at the outset that the empirical model under study respects some conventions that make the OLS estimation procedure valid regarding the specification, the behaviour of the residuals, the stability of the model, and so forth. Since such problems can yield spurious inferences and thus invalidate the results one obtains, it is increasingly common to test for possible departures from this set of conventions. Only when a researcher is reasonably satisfied that the empirical model provides a good statistical representation of the data, he or she should go on with an economic assessment of the model. For instance, suppose one wishes to test whether the income elasticity of money demand is equal to one in such an equation. The conclusion, he or she may reach may not be appropriate if important regressors, say for instance the financial transaction costs, have been omitted. Indeed, it is likely that his or her coefficient estimate of the income elasticity will then be biased.

Many tools for assessing the statistical adequacy of a regression equation have been devised in the last 20 years and are becoming easily accessible on basic econometric softwares. At the same time, there is often a unifying principle underlying these diagnostic tests that can help one to make use and to interpret what comes out of these statistics. In the next section, we will review both the general linear model and its conventions, the econometric problems resulting from the violations of these conventions, and some frequently employed estimators. Subsequently, we will present a general principle and a battery of diagnostic tests. Finally, we will discuss the fundamental philosophy of the methodology underlying modern econometrics.

2.2 *The General Linear Model, its Conventions and Some Commonly Used Estimators.*

2.2.1 Review of the general linear model, its conventions, and the OLS estimator.

Even without using rigourous derivations, it is first useful to remind ourselves of the general linear model, its conventions and the fundamental properties of the most basic and common estimator, i.e. the Ordinary Least Squares method.[1] This chapter cannot substitute for a formal presentation found in many textbooks, but it seeks to review important concepts and properties upon which are based most discussions in econometric methods.

Consider the following general linear model:

$$y_t = \mathbf{x}_t\beta + u_t \, , \qquad \forall \, t = 1, 2, \ldots, T \qquad (2.1)$$

where y_t, $\mathbf{x}_t$, β, and u_t are respectively, 1×1, $1 \times k$, $k \times 1$ and 1×1. The row vector $\mathbf{x}_t$ is defined so that a constant is included as a regressor. Equivalently, this can be expressed in matrix form as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \qquad (2.2)$$

where $\mathbf{y}$, $\mathbf{X}$, β, and $\mathbf{u}$ are matrices (or vector) of dimensions $T \times 1$, $T \times k$, $k \times 1$, and $T \times 1$.

The following set of conventions is also assumed:

1. Equation (2.2) constitutes the true model, and the vector of parameters β has finite values. $\mathbf{X}$ is a matrix made of the realizations of non degenerate stochastic variables, observed without measurement errors. The quantity $E(\mathbf{X}'\mathbf{X})^{-1}$ exists and is positive definite. Finally, the conditional expectation of $\mathbf{y}$ is linear in $\mathbf{X}$ only, i.e. $E(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\beta$.
2. The variables making up $\mathbf{X}$ are not correlated with the error term, i.e. $E(\mathbf{X}'\mathbf{u}) = 0$, which implies that $E(\mathbf{u}|\mathbf{X}) = 0$.
3. (i) The error term is not serially correlated, i.e. $E(u_t u_s|\mathbf{X}) = 0$ for all $t \neq s$.
   (ii) The conditional variance of the error term is constant, i.e. $E(u_t^2 \mid \mathbf{X}) = \sigma^2$.
4. The error term's distribution follows a normal distribution, i.e. $\mathbf{u} \mid \mathbf{X} \sim N(0, \sigma^2\mathbf{I})$.

---

[1] Stigler's (1981) quote reported in Maddala (1992) gives a good characterization of the special relationship between the OLS estimator and the applied economists. *"The method of least squares is the automobile of modern statistical analysis; despite its limitations, occasional accidents, and incidental pollution, it and its numerous variations, extensions and related conveyances carry the bulk of statistical analysis, and are known and valued by all."*

The OLS estimator $\hat{\beta}$ of the coefficients are obtained from minimizing the sum of squared residuals, i.e. solving the following problem

$$\min_{\hat{\beta}} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) = \min_{\hat{\beta}} \sum_{t=1}^{T} u_t^2 = \min_{\hat{\beta}} \sum_{t=1}^{T} (y - x_t \beta)^2 ,$$

which yields $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. It can also be shown that the least squares estimator of the variance of $\hat{\beta}$ is given by $\widehat{Var}\,(\hat{\beta}) = s^2\,(\mathbf{X}'\mathbf{X})^{-1}$, where $s^2 = \hat{\mathbf{u}}'\,\hat{\mathbf{u}}/(T-k)$.

As it is shown in every basic econometric textbook, the conventions listed under number 1 to 3 imply that the OLS estimator is the best linear unbiased estimator (B.L.U.E.). That is, not only $\hat{\beta}$ is an unbiased linear estimator in repeated samples but it is also the one with the smallest variance amongst any linear unbiased estimators. This is what is known as the Gauss-Markov Theorem. Under conventions 1 to 4, we also have: $\hat{\beta} \mid \mathbf{X} \sim N\,(\beta, \sigma^2\,(\mathbf{X}'\mathbf{X})^{-1})$. Finally, under conventions 1 to 3, provided some regularity conditions and the stationarity of the regressors, central limit theorems imply also the following asymptotic distribution:[2] $\sqrt{T}\,(\hat{\beta} - \beta) \overset{a}{\sim} N(0, \sigma^2 \lim_{T\to\infty} [E(\mathbf{X}'\mathbf{X}/T)]^{-1})$. We will now discuss in turn the role of conventions 1 to 4 and the problems one encounters if some of these assumptions are violated.

2.2.2 Review of the problems occurring from the violations of the underlying conventions and alternative estimators.

Typically, a least-squares estimator attempts to disentangle the portion of the variation in a dependant variable that can be attributed to the variation in each regressor separately. That is why any common variation between two regressors that cannot be dissociated from one another is mechanically discarded at the estimation stage.

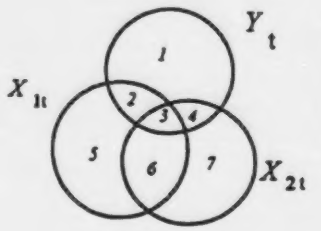


FIGURE 2.1

Following Kennedy (1992), this is illustrated in Figure 2.1 with a Venn diagram for a regression equation such as:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t, \qquad (2.3)$$

where $u_t \sim$ i.i.d. $N(0, \sigma^2)$. The circle associated with Y_t represents the variation in the dependent variable that we attempt to explain by the variation in X_{1t} and in X_{2t}. To estimate β_1, we want to use the information represented by area *2* only, which is common to both Y_t and X_{1t} alone. In the same way, only the information common to Y_t and X_{2t} alone is used to estimate β_2 (i.e. area *4* only). Since X_{1t} and X_{2t} are correlated, as represented by the intersection areas *3* and *6*, it is not possible to disentangle what portion of that area 3 comes from X_{1t} or from X_{2t}. This is why the larger is area *3*, the less information there is available to estimate β_1 and β_2. Consequently, the less precise will be $\hat{\beta}_1$ and $\hat{\beta}_2$. Moreover, the size of area *1* reflects the magnitude of σ^2.

---

[2] Dhrymes (1974) and White (1984) provide detailed presentations of various central limit theorems and other statistical theorems that can be used to derive asymptotic properties of estimators. Stationarity of a stochastic variable requires its

2.2.2.1 Violation of conventions 1 and 2.

2.2.2.1.1 Misspecifications and simultaneity bias.

Let us suppose that the Venn diagram in Figure 2.1 represents the true relation between Y_t, X_{1t} and X_{2t}. However, suppose that our empirical model is estimated while wrongly omitting the regressor X_{2t}:

$$Y_t = \beta_0 + \beta_1 X_{1t} + v_t, \tag{2.4}$$

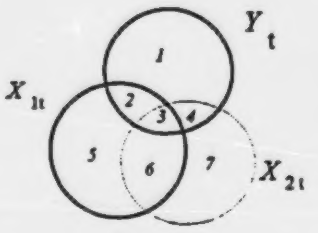


FIGURE 2.2

with $v_t = \beta_2 X_{2t} + u_t$. In this case, the least-square estimate of β_1 will be computed using all the information contained in areas *2* and *3*, as shown in Figure 2.2. Since area *3* should have been discarded because we cannot know whether it belongs to the explanatory power of X_{2t} on Y_t or to the explanatory power of X_{1t} to Y_t, the estimator of β_1 is generally biased. Also, since the estimate of σ^2 will be computed based on both areas *1* and *4*, it will be overestimated. Let us notice however that if X_{1t} and X_{2t} are uncorrelated, i.e. areas *3* and *6* do not exist, the estimator of β_1 is not biased. Nonetheless, the estimator of its variance-covariance matrix is biased upward since an area such as *4* is added up to area *1*, which gives a higher estimate of σ^2. Hence, the omission of an important variable implies that the OLS estimator is no longer B.L.U.E.. Formally, this type of misspecification amounts to a violation of the second convention presented in the previous section.

---

moments at different order to be independent of time. Chapter 3 and 4 will address further the implications related to the issue of stationarity.

In another case, suppose that irrelevant explanatory variables have been included in the regression (e.g. the true value of β_2 is zero), the least-squares estimator of β_1 and of its variance will not be biased. Referring to Figure 2.2, the wrong inclusion of X_{2t} will mean that the estimate of β_1 will be computed using only the area *2*, so that some information from area *3* will unfortunately be discarded. This loss of information will increase the variance-covariance matrix of the least-squares estimate of β_1. It does not lead however to the misguided use of irrelevant information, so that no conventions underlying the properties of the least-squares estimator are violated. The OLS estimator remains B.L.U.E. in this situation.

Violation of conventions 1 and 2 arises in three instances: (i) if the regressors represented by the X matrix suffer from measurement errors correlated with the error term **u**; (ii) if the lagged dependent variable is included amongst the regressors and there is serial correlation in the error term; or (iii) if there is simultaneity between some of the regressors and the dependent variable. Under all three scenarios, the error term will be correlated with the regressors, so that $E(\mathrm{X'u}) \neq 0$, and the OLS estimator of β is biased. For instance, considering that the true model is given by:

$$Y_t = \beta_0 + \beta_1 X_{1t} + u_t, \tag{2.5}$$

where $E(X_{1t}{}' u_t) \neq 0$. Referring to Figure 2.3, area *1*, which represents the influence of the error term, must intersect with the variation in X_{1t}. This is what is depicted by area *2*. Hence, the variation in Y_t captured in area *2* is both due to the influence of the error term u_t and that of the regressor X_{1t} and it is not possible to disentangle which is which. The regression of Y_t on an intercept and X_{1t} will therefore wrongly use all the information represented by areas *2* and *3* to compute the OLS estimate of β_1. That is why it will be biased, so that we need to use an estimator such as an instrumental variable (IV) estimator that possesses interesting statistical properties in large enough samples under some conditions.[3]

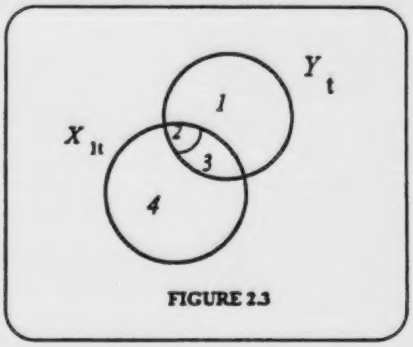


FIGURE 2.3

---

[3] The statistical justification of an IV estimator will rely upon asymptotic properties, such as convergence in probability, and asymptotic efficiency rather than the unbiasedness and the smallest variance properties in repeated samples. In any case, with time series data, since history unfolded only once, we are limited to work with only one sample.

2.2.2.1.2 Instrumental variable estimators.

Since $E(\mathbf{X}'\mathbf{u}) \neq 0$ and plim $T^{-1}\mathbf{X}'\mathbf{u} \neq 0$ for some elements of **X** (i.e. there exists contemporaneous correlation between the error term and some of the regressors) a solution to this problem would be to find a matrix of instruments **W** such that plim $T^{-1}\mathbf{W}'\mathbf{u} = 0$. Each row in **W** would be made of a vector $\mathbf{w}_t'$ of k elements, such that (1) either $E(\mathbf{w}_t'u_t) = 0$, or $\mathbf{w}_t$ and u_t would be independent at least asymptotically, and (2) $\mathbf{w}_t'$ would be highly correlated with the regressors $\mathbf{x}_t'$. Notice that for the columns of $\mathbf{x}_t'$ that are not correlated with u_t, the specific regressor in these columns is the best instrument for itself. Then, the instrumental variable estimator of β is derived from the sample counterpart of the expectation condition: $E(\mathbf{w}_t'u_t) \equiv E(\mathbf{w}_t'(y_t - \mathbf{x}_t'\beta)) = 0$, i.e. $(1/T)\,\mathbf{W}'(\mathbf{y} - \mathbf{X}'\beta) = 0$, so that:[4]

$$\hat{\beta}_{IV} = (\mathbf{W}'\mathbf{X})^{-1}\mathbf{W}'\mathbf{y} \quad , \tag{2.6a}$$

and $$\widehat{Var}(\hat{\beta}_{IV}) = \hat{\sigma}^2_{IV}(\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X})^{-1} = \hat{\sigma}^2_{IV}(\mathbf{W}'\mathbf{X})^{-1}\mathbf{W}'\mathbf{W}(\mathbf{X}'\mathbf{W})^{-1} \quad , \tag{2.6b}$$

where $\hat{\sigma}^2_{IV} = \dfrac{\hat{u}'_{IV}\hat{u}_{IV}}{T-k}$ and $\hat{u}_{IV} = \mathbf{y} - \mathbf{X}\,\hat{\beta}_{IV}$. Notice that the right-hand-side of equation (2.6b) is correct because we suppose that matrices **W** and **X** have the same dimension.

It is possible that a researcher has at one's disposal more instruments than regressors, say $s > k$. Even though the number of instruments could be larger than the number of regressors in the above formulae, this is not the best way to combine the instruments. One drawback of the simple IV estimator is that even though it is consistent and asymptotically normally distributed, it will tend to have a larger variance-covariance matrix than that of the OLS estimator, the smaller is the correlation between **X** and **W**. That is why an applied researcher will rather use a generalized instrumental variable estimator (GIVE) when confronting a problem of correlation between the regressors and the error term.

The GIVE procedure may be thought as the solution of the following problem:

$$\min_{\hat{\beta}_{GIVE}} Q_{GIVE}(\beta) \equiv (\mathbf{y} - \mathbf{X}\beta)'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{X}\beta)$$

which yields:

$$\hat{\beta}_{GIVE} = (\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}, \tag{2.7a}$$

---

[4] The simple IV estimator can be derived also as the solution to the minimization of $Q_{IV}(\beta) = (\mathbf{y} - \mathbf{X}\hat{\beta})'\,\mathbf{W}\,\mathbf{W}'\,(\mathbf{y} - \mathbf{X}\hat{\beta})$.

and $\hat{Var}(\hat{\beta}_{GIVE}) = \hat{\sigma}^2_{GIVE}(X'W(W'W)^{-1}W'X)^{-1}$ where $\hat{\sigma}^2_{GIVE} = \frac{\hat{u}'_{GIVE}\hat{u}_{GIVE}}{T-k}$. (2.7b)

Intuitively, the GIVE procedure aims at more efficiently combining the menu of instruments to maximize the correlation between weighted averages of the instruments and each of the regressors. In effect, it amounts in using $\hat{X}$ as instruments for X, which are defined as the predicted value of X obtained from the regression of X on W, i.e. $\hat{X} = W(W'W)^{-1}W'X$. Referring to these instruments, it is easy to see that the GIVE of β is a generalization of the IV estimator of β: $\hat{\beta}_{GIVE} = (\hat{X}'X)^{-1}\hat{X}'y$, where each column $\hat{X}$ is respectively used as an instrument for each column of X.[5] To assess the quality of the instruments, one can check the adjusted R^2 values in the OLS regressions of each column of X on W. It would also be useful to verify the significance level of an F test for the restriction that all coefficients other than the intercept are zero in an OLS regression of the GIVE residuals on W.

Notice that the GIVE minimand is the minimized value of the function $Q_{GIVE}(\cdot)$ evaluated at the values of the GIVE estimates:[6]

$$Q_{GIVE}(\hat{\beta}_{GIVE}) \equiv (y - X\hat{\beta}_{GIVE})'W(W'W)^{-1}W'(y - X\hat{\beta}_{GIVE}), \tag{2.8a}$$

$$= \hat{u}_{GIVE}{}'W(W'W)^{-1}W'\hat{u}_{GIVE}. \tag{2.8b}$$

This quantity will take a positive value except when the number of regressors is exactly equal to the number of instruments so that $Q_{GIVE}(\hat{\beta}_{GIVE}) = 0$, i.e. if $s = k$. As we will discuss later, this quantity can be employed to construct some statistical tests.

2.2.2.2 Violation of convention 3.

2.2.2.2.1 Nonspherical error term.

Convention 3 can be restated as $E(uu') = \sigma^2 I_{T \times T}$. We then say that the variance-covariance matrix of the error term is spherical. This will be violated in any of the following circumstances: (i) if there is serial correlation in the error term; or, (ii) if the variance of the error term is not constant. Then, $E(uu') = \Omega$: (i) with non zero off-diagonal elements if there is serial correlation, and (ii) with non-constant diagonal elements if there is heteroskedasticity in the error term. Either of these problems do not lead *per se* to violations of conventions 1 and 2, so that the OLS estimator of β remains unbiased. Intuitively, the unbiasedness of $\hat{\beta}$ is derived from the fact that in repeated samples even if the error terms specific to each sample are not independently distributed or have a time-varying variance, they are not correlated with the regressors on average. However, the OLS estimator of the variance-covariance matrix of $\hat{\beta}$, that is computed as $\hat{Var}(\hat{\beta}) = s^2 (X'X)^{-1}$, is itself biased since the true variance of the OLS estimator can be shown to be equal to $Var(\hat{\beta}) = (X'X)^{-1} (X' \Omega X)(X'X)^{-1}$. Therefore, this renders inefficient the OLS estimator and incorrect any inference based on the values reckoned this way. There are several solutions to either the serial correlation problem or the heteroskedasticity problem.

---

[5] Notice that the R^2 reported by many computer softwares under an IV or 2SLS option does not have the same interpretation as with the OLS procedure. One alternative proposed by Maddala (1992) is to compute the squared correlation between y and its GIVE fitted value. Interested readers are referred to Maddala (1992) for other issues pertaining to goodness-of-fit measures with time-series data.

[6] A generalisation of the GIVE procedure is the Generalised Method of Moments (GMM) estimator which can be employed to estimate a simultaneous set of non-linear equations by instrumental variables. This technique has become increasingly popular in macroeconometrics to estimate Euler orthogonality conditions. For instance, these can derived from the first-order conditions of an economic agent's optimisation problem under rational expectations. See Hansen (1982) for a presentation of the theory of GMM, as well as Hansen and Singleton (1982) for an application. Greene (1990), and Davidson and MacKinnon (1993) provide textbook expositions.

2.2.2.2.2 Feasible generalized estimator and heteroskedasticity and serial correlation consistent estimator of the variance-covariance matrix of the regression coefficients.

If one knows the underlying form of serial correlation or of heteroskedasticity, it is possible to get a consistent estimator of Ω. An estimate $\hat{\Omega}$ can be computed using the postulated functional form of these deviations from spherical stochastic disturbances, and a Feasible Generalized Least-Squares (FGLS) estimator can be evaluated. Namely,

$$\min_{\hat{\beta}_{FGLS}} (y - X\beta)' \hat{\Omega}^{-1} (y - X\beta)$$

would give:

$$\hat{\beta}_{FGLS} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y \quad \text{and} \quad \hat{Var}(\hat{\beta}_{FGLS}) = (X' \hat{\Omega}^{-1} X)^{-1}. \tag{2.9}$$

The FGLS estimator can be shown to be consistent and asymptotically efficient.[7, 8]

The FGLS estimator is not a useful option if we cannot postulate the underlying form of nonsphericalness of the disturbances so that we eliminate the problem. Since this problem does not invalidate the unbiasedness (and consistency) of the OLS estimator of β, an alternative would be to construct a different estimator of $Var(\hat{\beta})$ which would be consistent in presence of

---

[7] Technically, since the estimator of Ω is a function of y_t, the FGLS estimator of β is not linear. Furthermore, it is also biased in repeated samples. That is why we then rely on its desirable asymptotic properties. Only, in the case where we would know Ω up to a factor of proportionality, then the GLS estimator would be B.L.U.E.

[8] Alternatively, in some situations (such as when the error terms follow an exact AR(1), AR(2) or MA(q) process, or some specific form of heteroskedasticity) one can also estimate the model by exact maximum likelihood method if assuming joint-normal disturbances. This estimator will possess the useful asymptotic properties.

heteroskedasticity and/or serial correlation. White (1980) and MacKinnon and White (1985) have proposed to use the following heteroskedasticity-consistent (HC) estimator of the variance-covariance matrix of the regression coefficients:

$$\hat{Var}_{HC}(\hat{\beta}) = \left(\tfrac{T}{T-k}\right) (X'X)^{-1}\, \hat{S}^{*}\, (X'X)^{-1} \tag{2.10}$$

where $S^{*} = X'\,\Omega^{*}X = \sum_{t=1}^{T} u_t^2\, \mathbf{x}_t \mathbf{x}_t' = X'\,\mathrm{diag}\left[u_t^2\right]_{T\times T} X$. The factor $(T\,/\,T\text{-}k)$ adjusts the estimator for the number of degrees of freedom.

Following the same logic, Newey and West (1987) have proposed a heteroskedasticity and autocorrelation consistent (HSCC) estimator of the variance-covariance matrix of the regression parameters when the serial correlation follows an unspecified structure:

$$\hat{Var}_{HSCC}(\hat{\beta}) = \left(\tfrac{T}{T-k}\right) (X'X)^{-1}\, \hat{S}^{**} (X'X)^{-1} \tag{2.11}$$

where $\hat{S}^{**} = X'\,\Omega^{**}X = \hat{S}^{*} + \sum_{j=1}^{m} w_{jm} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j} (\mathbf{x}_t \mathbf{x}_{t-j}' + \mathbf{x}_{t-j} \mathbf{x}_t')$. Essentially, this amounts to finding a consistent estimate of the unknown variance-covariance matrix Ω of the error-term whose elements on the diagonal are made of the squared OLS residuals while the off-diagonal elements are set as weighted cross-products of these residuals at different lags. The w_{jm} weights serve to shrink the off-diagonal elements towards zero as they correspond to elements relatively far from the diagonal, i.e. the correlations between residuals for periods relatively far from one another.[9, 10]

Hence, inference on the estimate of β can be conducted with usual test statistics constructed with the OLS estimator and the associated HSCC or HC estimate of the variance-covariance matrix of $\hat{\beta}$.[11]

---

[9] If the regression model is known to exhibit moving average errors of known and finite order *m*, such as in the case of overlapping data for which the forecast horizon exceeds the sampling interval, the uniform window with $w_{jm} = 1$, for all *j* is appropriate. In other cases, there are better alternatives. Furthermore, if the chosen value of *m* is large relative to the number of observations, the estimate of the variance matrix with uniform weights may be negative-definite. However, the Bartlett window, as originally used by Newey and West (1987), and the Parzen window ensure a positive-definite variance matrix. The Bartlett window has weights $w_{jm} = 1 - (j\,/\,(m+1))$, for $j = 1, 2, \ldots, m$. The Parzen window has weights defined as $w_{jm} = 1 - 6\,(j/(m+1))^2 + 6\,(j/(m+1))^3$, for $j \in [1, (m+1)/2]$, and $2\,(1 - j/(m+1))^3$, for $j \in \,](m+1)/2, m]$. Pesaran and Pesaran (1991) suggest that the Parzen window is generally preferable.

[10] One practical problem with the construction of the HSCC estimator is the determination of the lag window *m* needed to define the weights. Pesaran and Pesaran (1991) recommend a value of *m* not to be larger than 1/3 of the sample size. Recently, Andrews (1991) and Andrews and Monahan (1992) have proposed other HSCC estimators of the variance matrix with a data-dependent method for selecting *m*.

[11] An estimator of the HC and of the HSCC variance-covariance matrix exists also for the regression coefficients

It is important to emphasize however that often a behaviour of the residuals that may appear to be significantly symptomatic of serial correlation reflects missing explanatory variables capturing statistically the dynamics of adjustment of the dependent variable. It may also happen that symptoms of heteroskedasticity are caused by an improper functional form or omitted explanatory variables. In these cases, the automatic use of a generalized least-squares estimator or of a HSCC estimator will not resolve the true misspecification problem and will not eliminate any biases or inconsistencies in the estimates. One must then give further thoughts to the specification he or she wants to estimate.

2.2.2.3 Violation of convention 4: non-normal error term.

Let us define the j^{th} moment of random variable as $\mu_j = E[(u_t - E(u_t))^j]$. If the error terms u_t follow a normal distribution $N(0, \sigma^2)$, then $\mu_1 = 0$, $\mu_2 = \sigma^2$, $\mu_3 = 0$, and $\mu_4 = 3\ (\sigma^2)^2$. The first and second moments provide a characterization of the location and of the dispersion of the distribution, respectively. The third moment means that the distribution of a normal error term is symmetric about its mean, while the fourth moment implies that the tails of the distribution tend to die off, i.e. they are not too fat.

Normality is not necessary to derive that the OLS estimator is B.L.U.E. under conventions 1 to 3. It is however helpful if one wants to know the distribution of the estimator of β and subsequently construct test statistics with a known distribution in finite samples. The derivation of the Maximum Likelihood estimator (MLE) of β on the other hand requires that we specify the distribution of the underlying stochastic disturbances of the model. As discussed in Pagan and Hall (1983), by referring to a Central Limit Theorem, the OLS estimator is asymptotically consistent and normally distributed even though the residuals in the sample do not appear to follow a normal (or Gaussian) distribution. The MLE can then be interpreted as quasi-MLE. What is required for the asymptotic properties of the OLS estimator to hold is that the error terms be identically and independently distributed and that the fourth moment of the disturbances be finite.

---

estimated by a Generalized Instrumental Variable method. See Pesaran and Pesaran (1991) and Davidson and MacKinnon (1993).

2.3 Statistical Inadequacy and Specification Tests in Econometric Models.

2.3.1 Assessment of an empirical model by variable addition: a general principle.

The statistical adequacy of the general linear model can be summarised with three requirements. (1) The mean conditional upon the information set represented by the regressors is given by $E(\mathbf{y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$. (2) The conditional variance $E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})|\mathbf{X}]$ is a constant σ^2. (3) The conditional distribution of y is normal, so that the higher conditional moments of y are solely function of the conditional mean and variance.

Many of the inadequacies of an empirical model occurring from the violation of the basic conventions of the general linear model can be cast in terms of an omitted variable problem. This interpretation therefore suggests a natural way to construct misspecification tests.

2.3.1.1 Augmenting the conditional mean.

Let us first consider the problems related with the conditional mean of a variable y. We will first consider the non-structural case of a single equation. Let us suppose that the empirical model we are going to estimate is represented by equation (2.2), which we write again for convenience:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \tag{2.12}$$

Let us suppose furthermore that an alternative specification of equation (2.12) would be given by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e} \ , \tag{2.13}$$

where $\mathbf{Z}$ is of dimension $T \times q$, and γ is $q \times 1$. Unless the true value of γ is zero, model (2.12) is therefore misspecified. Of course, if we knew *a priori* what Z could be, we would only need to begin our search for an adequate specification with model (2.13) (that is more general) and test the null hypothesis H_0: $\gamma = 0$ with a t- or F-test.[12] Another possibility would be to use some proxy variables $\tilde{\mathbf{Z}}$ in place of Z, so that we can perform the test of H_0: $\gamma = 0$ once the following equation is estimated:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \ , \tag{2.13'}$$

---

[12] Notice that if lagged values of the dependent variables are included amongst the x_i's, the t- and F- tests are only asymptotically valid, even when the errors are normally distributed.

where $\varepsilon = \left[e + (Z - \tilde{Z})\gamma)\right]$. Under the null hypothesis that γ is equal to zero, then model (2.13) is statistically adequate in its conditional mean, and equation (2.13) and (2.13') are identical. Moreover, the higher the correlation between Z and $\tilde{Z}$, the higher the power of diagnostic tests based on equation (2.13') will be, *ceteris paribus*.[13]

It is possible to apply a test of variable addition for H_0: $\gamma = 0$ by constructing a Wald (W, or equivalently, an F-test), a Likelihood ratio (LR), or a Lagrange multiplier (LM) statistics based on representation (3.2'). While these statistics will be asymptotically equivalent, their conclusions may however differ in small sample.[14]

In the case of the OLS estimator, using the *F*-variant of the test amounts to:[15]

1. First stage regression: regress y on X, and get the residual sum of squares of this regression, i.e. SSE^R. This is associated with the restricted model.
2. Auxiliary regression, associated with the unrestricted model: regress y on X and $\tilde{Z}$, and get the residual sum of squares, i.e. SSE^U.
3. To test H_0: $\gamma = 0$ with a statistic that asymptotically follows an *F*-distribution with *q* and *T-k* degrees of freedom, calculate $F = \dfrac{(SSE^R - SSE^U)/q}{SSE^U/(T-k)} \overset{a}{\sim} F(q, T-k)$.

An easy way to construct the LM variant of the test for an OLS regression can be summarized as follow:

1. First stage regression: regress y on X to get the OLS estimate $\hat{\beta}$.
2. Compute the OLS residuals $\hat{u} = y - X\hat{\beta}$.
3. Auxiliary regression: estimate $\hat{u} = X\,b + \tilde{Z}\,\gamma + v$ with a sample of T_a observations and get its unadjusted R^2.
4. To test H_0: $\gamma = 0$ with a LM statistic that asymptotically follows a χ^2-distribution with *q* degrees of freedom, calculate $LM = T_a \cdot R^2 \overset{a}{\sim} \chi^2(q)$.

---

[13] As discussed by Pagan (1984), the power of the test will also be function of the nature of X.

[14] Berndt and Savin (1977) have shown that the following inequality $LM \leq LR \leq W$ holds in small sample for the linear regression model, when testing linear restrictions on the conditional mean parameters β. In practice, this may suggest that at times a smaller level of significance than usual could be warranted for a W-statistic in a small sample.

[15] For linear restrictions imposed on a model estimated by OLS, the *F*-statistics with OLS variance-covariance matrix can be constructed by the formula shown in step 3. In other cases (such as with the GIVE procedure or with an HC or an HSCC estimator of the variance-covariance matrix), the *F* or Wald statistics, to test linear restrictions expressed as H_0: $R\,\beta = r$ (where R is a known matrix of dimension $q \times k$ and r is $q \times 1$ vector must be constructed using the general expression: $W = (R\,\hat{\beta} - r)'[R\,\widehat{\text{var}}(\hat{\beta})R']^{1}\,(R\hat{\beta} - r)$. Notice that the Wald statistic is related to the *F*-statistic, as $W = q\,F$.

The tests we have just presented are appropriate provided that the error terms are homoskedastic and independent. It is possible to compute versions of the tests that use the HC estimator or the HSCC estimator of the variance-covariance matrix, but often this will require programming with linear algebra unless the option is pre-programmed in the software being used.[16] To construct heteroskedasticity-robust misspecification tests, Davidson and MacKinnon (1993) discuss also the following procedure.

1. First stage regression: regress y on X, and get the vector of residuals $\hat{\mathbf{u}} = [\hat{u}_1, ..., \hat{u}_T]'$.
2. Compute the q vectors of residuals $\hat{\varsigma}_1, \hat{\varsigma}_2, ..., \hat{\varsigma}_q$ obtained from the regression of each variables in $\tilde{\mathbf{Z}}$ on X.
3. Calulate a set of q auxiliary regressors ς_i by multiplying the t^{th} element of each vector $\hat{\varsigma}_i$ by $\hat{u}_t$ for $i = 1, ..., q$.
4. Auxiliary regression: for a sample of T_a observation, estimate $\iota = [\tilde{\varsigma}_1, \tilde{\varsigma}_2, ..., \tilde{\varsigma}_q]\gamma + v$, where ι is a vector of 1s, and get the residual sum of squares of this regression, i.e. SSE.
5. To test H_0: $\gamma = 0$ with a LM statistic that asymptotically follows a χ^2-distribution with q degrees of freedom, calculate $LM = (T_a - SSE) \sim \chi^2(q)$.

Under the heading of augmenting the conditional mean, we can regroup several important diagnostic tests, namely for omitted variables, for functional form, for stability, for serial correlation, and for exogeneity. Essentially, what makes these tests different from one another is the definition of the proxy variable $\tilde{\mathbf{Z}}$, which depends on the hypothesis one addresses. Table 2.1 presents different tests that are discussed in the following subsections.

---

[16] Footnote 15 shows the formula that needs to be evaluated. However, softwares, such as EViews 1.0, do the necessary manipulations automatically when invoking the HC covariance matrix.

Table 2.1.- Inadequacy diagnostic tests - Augmenting the conditional mean.

From $\hat{u} = y - X\hat{\beta}$ and $\hat{\beta} = (X'X)^{-1}X'y$, estimate $\hat{u} = Xb + \tilde{Z}\gamma + v$ and test $H_0 : \gamma = 0$ using $LM = T_a \cdot R^2 \overset{a}{=} \chi^2(q)$

Name of the test	Nature of the problem	Definition of $\tilde{Z}$	T_a	q
RESET test	Specification test for omitted variables Inadequacy of the functional form	$\tilde{Z} = [\hat{y}^2, \hat{y}^3, \hat{y}^4, \ldots, \hat{y}^L]$ where $\hat{y} = X\hat{\beta}$	T	$L\text{-}1$
Chow test	Test for the constancy of the coefficients. The first subsample has T_1 observations while the second subsample has T_2 observations.	$\tilde{Z} = [dx_1, \ldots dx_k]$ $d_t = 0$ for $t \leq T_1$ $= 1$ for $T_1 < t \leq T_1 + T_2$	$T = T_1 + T_2$	k
One-step ahead prediction error test	To assess the stability of the model. The first subsample has T_1 observations while the second subsample has T_2 observations.	$\tilde{Z} = \begin{bmatrix} 0_{T_1 \times T_2} \\ I_{T_2 \times T_2} \end{bmatrix}$	$T = T_1 + T_2$	T_2
Autocorrelation	To assess the evidence of uncorrelated error terms.	$\tilde{Z} = [\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-q}]$	$T - q$	q
Hausman test	To test for the exogeneity of the regressors. Let W be a matrix of q instruments such that $\hat{X} = E[\lvert X \rvert W] = W\hat{\delta}$ where $\hat{\delta} = (W'W)^{-1}W'X$	$\tilde{Z} = W\hat{\delta} - X = \hat{X} - X$	T	q
Plosser, Schwert, and White differencing test	General test of model misspecification	If $y_{t-1} \not\subset X$, $\tilde{Z} = X_{t+1} + X_{t-1}$	$T\text{-}2$	k
		If $y_{t-1} \subset X_t$, $\tilde{Z} = [y_{t-2} \vdots (X_{t+1} + X_{t-1})]$	$T\text{-}2$	k

2.3.1.1.1 The RESET test.

Ramsey (1969) devised a test for omitted variables (or specification error), that is based on the hypothesis that missing regressors might be correlated with the powers of the fitted values of the dependent variable, i.e. $\hat{y}^j = (X\hat{\beta})^j$, for $j = 2, \ldots, L$. Monte-Carlo experiments conducted by Ramsey and Gilbert (1972) suggest that setting $L = 4$ provides a fair enough degree of power. In practice, however, an applied researcher may find that sometimes there is strong multicollinearity between $\hat{y}^2$, $\hat{y}^3$ and $\hat{y}^4$, so that the matrix inversion required to conduct the estimation of the auxiliary regression cannot be computed. We then suggest first to conduct the test with $L = 3$. If there is still a near perfect collinearity problem, then one should settle for $L = 2$.

Rejection of the null hypothesis by this test has also another interpretation. Indeed, the RESET test may reveal evidence of departure from linearity. This is intuitively understandable, if one thinks of approximating a general nonlinear function of X by a Taylor series expansion. The maintained model would correspond to the first term of the expansion, while departures from nonlinearity could be revealed by the statistical significance of the powers of $\hat{y}$, which could be correlated with the higher order terms of the expansion.

The RESET test can be generalized to the case of nonspherical errors due to either some unknown form of heteroskedasticity or serial correlation. In these cases, it is suggested to construct the Wald version of the omitted variable test using a HC or HSCC variance-covariance matrix of the OLS estimate of γ. When there is evidence of heteroskedasticy, Davidson and MacKinnon's (1993) procedure outlined previously can also be used. Notice however that evidence of serial correlation in dynamic models may often suggest a problem of missing regressors required to capture the dynamics of the model, so that the estimates of the coefficients are themselves biased and inconsistent.

If the RESET test provides evidence of an econometric problem, there is no solution but to give more thoughts to the empirical formulation of the model !

2.3.1.1.2 The Chow test for constancy of the coefficients.[17]

If the regression coefficients vary discretely at specified time points, the model is not stable. Under the maintained hypothesis that the variances of the errors are equal over all subperiods, a Chow test of the stability of the regresion coefficients can be easily constructed as a test of variable addition. It suffices to test the significance of a set of variables that are defined as the element by element product of each regressor with an indicator variable. To test for one shift in the regression coefficients over the whole sample, the dummy variable is defined as being zero for the first subsample, and one for the second subsample.

If errors are heteroskedastic, the test can be applied with its Wald version using the HC variance-covariance matrix or with Davidson and MacKinnon's (1993) procedure.

---

[17] There exist many variants of stability and predictive failure tests. See Pesaran, Smith, and Yeo (1985) for a survey.

2.3.1.1.3 The predictive failure test.[18]

If a model is stable, it should be capable of adequately predicting observations in a *post*-estimation sample of data. For instance, the model may have been initally estimated on T_1 observations, and we would want to assess its predictive value over the next T_2 observations[19]. In this case, Pagan and Nicholls (1984) suggest to augment the conditional mean by a set of indicator variables d_j, which are set to zero for all observations but the $(T_1+j)^{th}$, where $j = 1, \ldots, T_2$. Hence, the columns of the matrix $\tilde{\mathbf{Z}}$ are made of the indicator variables. The vector γ of coefficients (of dimension $T_2 \times 1$) can be interpreted as the one-step ahead prediction error, i.e. $\{\hat{y}_{T_1+j} - y_{T_1+j}\}$.

If the error terms are heteroskedastic, then one can use the Wald version of the test with the HC variance-covariance matrix of the auxiliary regression coefficients, or the proceduree suggested by Davidson and MacKinnon (1993). However, if the error terms are serially correlated, Pagan and Nicholls (1984) have shown that the indicator variables need to be constructed somewhat differently.[20] Finally, they have also proposed another extension of their test for multi-step ahead prediction errors.

2.3.1.1.4 The serial correlation test.

Considering the general linear model, $y_t = \mathbf{x}_t' \beta + u_t$, if the error terms are serially correlated, this means that they either follow an order p autoregressive process, i.e. AR(p),

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + e_t, \quad e_t \sim i.i.d.(0, \sigma_e^2) \ , \tag{2.14}$$

an order q moving average process, i.e. MA(q),

$$u_t = e_t + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \ldots + \rho_q e_{t-q}, \quad e_t \sim i.i.d.(0, \sigma_e^2) \ , \tag{2.15}$$

or a mixed ARMA(p, q) process.

The Breusch (1978) and Godfrey (1978) test of serial correlation consists simply to add lagged values of the OLS residuals obtained from the first stage regression, then to test their significance. This test is certainly preferable to the often used Durbin-Watson statistics and its variants, as it is applicable even if the lagged dependent variable is included amongst the regressors.[21] Notice however that this test will only reveal the importance of serial correlation, not its precise nature.

---

[18] More precisely, as shown in Pagan and Nicholls (1984) this is a one-step ahead predictive failure test.

[19] For instance, the estimation could be performed on the first 90 per cent of the observations, while the rest of the data could be used for out-of-sample prediction evaluation.

[20] In terms of a notation that will be formally introduced in the next chapter, technically, if $u_t \sim$ AR(p), e.g. $\phi(L)\, u_t = e_t$, then its MA representation is given by: $u_t = [\phi(L)]^{-1} e_t = \psi(L)\, e_t$ where $\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \ldots$. Then, $d_{j,T1+k}$ is set equal to 1, for $k = j$; the estimate of ψ_{k-j} for $k \geq j$, $j = 1, \ldots, T_2$, and $k = 1, \ldots, T_2$. Otherwise, $d_{j,t} = 0$, for $t \leq T_1$. If $u_t \sim$ MA(q), then $d_{j,T1+k}$ would be constructed from the MA coefficients directly.

Finally, as it was mentionned earlier, evidence of serial correlation in dynamic models may suggest a problem of missing regressors required to capture the dynamics of the model. This suggests that before concluding that one is facing a true problem of serial correlation, he or she must feel very confident that the dynamic adjustment is adequately captured in a statistical sense, if not, the estimates that are obtained in ignoring this problem will be biased and inconsistent.

2.3.1.1.5 The Hausman test of exogeneity.

The Hausman (1978) test is a very general test of the null hypothesis that the regressors in **X** are contemporaneously uncorrelated with the error term **u**. In effect, under the null hypothesis that the specification is correct, estimates of β that are obtained either by OLS or GIVE will be consistent and statistically close to each other. If however, the $E(\mathbf{X}'\mathbf{u}) \neq 0$, then the OLS estimator should be inconsistent while the GIVE should be consistent. Hence, we would expect that the OLS estimates of β be statistically different from the GIVE estimates.

An omitted variable interpretation of the Hausman test that all the regressors are exogenous is described in Table 2.1.[22] Essentially, one needs to add $\left[\hat{\mathbf{X}} - \mathbf{X}\right]$ in an auxiliary regression and to test whether the associated coefficients are non-significant. The columns of $\left[\hat{\mathbf{X}} - \mathbf{X}\right]$ are the residuals from the OLS regression of each regressor on all the available instruments (or predetermined variables) **W**.

Notice that if some of the regressors can be unmistakenly taken to be exogenous or at least predetermined from the outset, then the test of exogeneity can be performed on the subset of regressors that might be correlated with the error term. For instance, suppose that the model to be estimated is:

---

[21] As we will discuss in the next chapter, it is also more powerful than portmanteau statistics such as the Quenouille tests and its variants.

[22] The original derivation of this misspecification test was cast in terms of a direct comparison between the GIVE and the OLS estimators. The programming of this format of the test requires that it be programmed in matrix algebra. Reinterpreting it as an omitted variable test means that it can be performed using any regression package with OLS.

$$\mathbf{y} = \mathbf{X}\,\beta + \mathbf{u} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \mathbf{u} \qquad (2.16)$$

where $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ are of dimensions $k_1 \times T$, $k_2 \times T$ and $k_3 \times T$, and $\mathbf{X}_3$ is made up of predetermined regressors. To construct the variables needed to apply the Hauman test on $\mathbf{X}_1$ and $\mathbf{X}_2$, one would regress each column of theses matrices on the instruments **W**, and keep the residuals from each regression. No such variables are needed for $\mathbf{X}_3$.

However, if one is interested in testing only a subset of $[\mathbf{X}_1 \;\; \mathbf{X}_2]$ for exogeneity, the test needs to be carried out differently. For instance, suppose that we assume that $\mathbf{X}_1$ is endogenous, but that we want to test that $\mathbf{X}_2$ is exogenous. We then need to compare the GIVE estimate of β_2 under the assumption of exogeneity of $\mathbf{X}_2$ with the GIVE estimate of β_2 under the assumption of $\mathbf{X}_2$ being endogenous. Following Spencer and Berk (1981), this can be done in the following manner:

1. Construct the k_1 variables $\check{\mathbf{X}}_1$ as the predicted values obtained from the regressions of each column of the matrix $\mathbf{X}_1$ on $\mathbf{X}_2$, and the instruments **W**, i.e. as if $\mathbf{X}_2$ were exogenous.
2. Construct the $(k_1 + k_2)$ variables $\tilde{\mathbf{Z}}_1 = \mathbf{X}_1 - \hat{\mathbf{X}}_1$ and $\tilde{\mathbf{Z}}_2 = \mathbf{X}_2 - \hat{\mathbf{X}}_2$ as the residuals from the regressions of each column of the matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ on the instruments **W**, i.e. as if $\mathbf{X}_2$ were endogenous.
3. Estimate the following equation by OLS:

$$\mathbf{y} = \check{\mathbf{X}}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{X}_3\beta_3 + \tilde{\mathbf{Z}}_1\gamma_1 + \tilde{\mathbf{Z}}_2\gamma_2 + \nu$$

and test the joint null hypothesis that $\gamma_1 = \gamma_2 = 0$ with an *F*-test.

Notice that the Hausman test is justified on asymptotic grounds.

2.3.1.1.6 Plosser, Schwert and White differencing test.

This test is an example of what is called a data transformation test. Under the null hypothesis of no misspecification, the least-square estimates obtained with the raw data should be statistically similar to those obtained from linearly transformed data. The Plosser, Schwert and White (1982) test compares the OLS estimates of the original model in levels with those obtained from the model in first-differences.[23] This differencing test is applicable provided that the regressors are stationary, and is robust against a broad range of alternative specifications that would lead to a correlation between the regressors and the error term (e.g. omitted variables, measurement error, simultaneity bias). These problems would render biased or inconsistent estimators of the regression coefficients. If the differencing test suggests a rejection of the null hypothesis, it does not identify exactly of the problem but it warns that one should not proceed with inference or prediction on the false basis that the original model was adequate.

---

[23] Let x_t be a variable in level, its first-difference is a linear transformation of x_t, given by $x_t - x_{t-1} = (1-L)\,x_t$. The lag

The differencing test can easily be constructed as an omitted variable test where we need to test the statistical significance of added variables $\tilde{\mathbf{Z}}$. As shown, in Table 2.1, these variables can be constructed as the sum of the one-period lag and the one-period lead of the original regressors. The only *caveat* is that if the lagged dependent variable y_{t-1} is originally included amongst the regressors, then the column of $\tilde{\mathbf{Z}}$ that is associated with it should be defined as y_{t-2}.

2.3.1.2 Augmenting the conditional variance.

The motivation behind different tests of heteroskedasticity is based on the idea that under the null hypothesis of homoskedastic disturbances, $E(u^2) = \sigma^2$ is time-invariant. This suggests that we augment an auxiliary regression for the squared residuals with some variables that could statistically mimic the source of heteroskedasticity, so that we would consider an equation such as $\hat{\mathbf{u}}^2 = \sigma^2 + \tilde{\mathbf{Z}}\gamma + \mathbf{v}$. Table 2.2 presents some variants of these tests.

Table 2.2- Inadequacy diagnostic tests - Augmenting the conditional variance.

From $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta}$, $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\iota' = [1, 1, \ldots 1]$, estimate $\hat{\mathbf{u}}^2 = \sigma^2\iota + \tilde{\mathbf{Z}}\gamma + \mathbf{v}$ and test $H_0 : \gamma = 0$ using $LM = T_a \cdot R^2 \overset{a}{=} \chi^2(q)$

Name of the test	Definition of $\tilde{\mathbf{Z}}$	T_a	q
Ramsey's test for heteroskedasticity	$\tilde{\mathbf{Z}} = [\hat{\mathbf{y}}, \hat{\mathbf{y}}^2, \hat{\mathbf{y}}^3, \ldots, \hat{\mathbf{y}}^L]$ where $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$	T	L
White's test for heteroskedasticity	$\tilde{\mathbf{Z}} = [x_1, \ldots, x_k, x_1^2, (x_1 x_2), \ldots (x_1 x_k), \ldots]$	T	$\frac{k(k+1)}{2} - 1$
	$\tilde{\mathbf{Z}} = [x_1, \ldots, x_k, x_1^2, \ldots x_k^2]$	T	$2k$
	$\tilde{\mathbf{Z}} = [x_1^2, (x_1 x_2), \ldots (x_1 x_k), \ldots, x_k^2]$	T	$\frac{k(k-1)}{2} - 1 - k$
Test for ARCH effect	$\tilde{\mathbf{Z}} = [\tilde{\mathbf{z}}_{t,}]$ where $\tilde{\mathbf{z}}_t = [\hat{u}_{t-1}^2, \hat{u}_{t-2}^2, \ldots \hat{u}_{t-q}^2]'$	T-2	q

If the heteroskedasticity problem is thought of as originating from a relationship between powers of the conditional mean of y and the variance of u, then $\tilde{\mathbf{Z}}$ can be defined as

---

operator L^j is defined as a tranformation of a variable that shifts the variable j periods backward : $L^j x_t = x_{t-j}$

some power of the least-squares predicted value of y. Generally, following Ramsey (1969) and Koenker (1981), this version of the test is conducted with the first or the second power of $\hat{y}$.

The White (1980) test for heteroskedasticity is intuitively based on a comparison of the OLS variance-covariance matrix of the parameters and the GLS estimator of this variance-covariance matrix. This is derived in postulating that the error variance is related to variables correlated with the regressors, their squares and their cross-products so that the OLS estimator of the variance-covariance matrix would be inconsistent. To apply this test, it suffices to augment the appropriate auxiliary regression based on the squared residuals with these variables. In practice, however, if k is a large number, the inclusion of all the regressors, the squares and the cross-products can lead to a situation where we would run out of observations to estimate the auxiliary regression. That is why some computer programs with a built-in variant of White's test do not include the cross-products while others choose to ignore the potential role of the regressors themselves. Notice finally that both Ramsey's and White's tests are robust to non-normal errors.

Finally, another test of heteroskedasticity that is often encounters, particularly when dealing with financial variables, is the test for autoregressive conditional heteroskedasticity (ARCH). It was observed by Engle (1982) that large and small forecast errors tend to occur in clusters so that the conditional variance of u_t may be an autoregrssive function of the size of the past errors. For instance, an ARCH(q) process would be written as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 + v_t. \tag{2.17}$$

This suggests that a test for ARCH(q) versus ARCH(0) can be built by adding lagged squared OLS residuals in the auxiliary regression of the squared residuals.[24]

2.3.1.3 Augmenting the higher conditional moments: a test of normality.

As discussed earlier, the normality of the error term is not always strictly required, since *"most results concerning the distributions of estimators and test statistics that are derived assuming normality remain asymptotically valid for many types of non-normality."* (See Godfrey (1988)). Nevertheless, it is a convenient assumption that is often taken for granted and it is useful to know better the distribution of the test statistics in finite samples, or to construct prediction intervals.[25]

---

[24] Bollerslev (1986) has also developed generalizations of the ARCH process, assuming that the conditional variance of the forecast errors can also be a function of past variances. That is a called a GARCH process. When this process is non-stationary, we also refer to IGARCH processes.

Under the normality of the underlying errors, we would expect that third centered moment of the error term be equal to zero and that the fourth centred moment be equal to 3 times the squared variance of the error term. Therefore, a test of normality of the residuals can be seen as testing the joint hypothesis that H_0: $\gamma_3 = \gamma_4 = 0$ in the following equation system:

$$\hat{u}_t^3 = \gamma_3 + v_{3t} \tag{2.18a}$$

$$\left[\hat{u}_t^4 - 3\hat{\sigma}^2\hat{u}_t^2\right] = \gamma_4 + v_{5t} \tag{2.18b}$$

where $\hat{u}_t$ is the OLS residual from the estimation of the empirical model, while v_{3t} and v_{5t} are error terms in the respective equations. For the case where an intercept is included in the empirical model, Bera and Jarque (1981) have shown that a LM test of this hypothesis can be constructed as:

$$LM_N = \frac{T}{6}\cdot\left[\frac{\hat{\mu}_3{}^2}{(\hat{\sigma}^2)^3} + \frac{1}{4}\cdot\left(\frac{\hat{\mu}_4}{(\hat{\sigma}^2)^2} - 3\right)^2\right] \overset{a}{\sim} \chi^2(2), \tag{2.19}$$

where $\hat{\sigma}^2 = \Sigma_{t=1}^T \hat{u}_t^2/T$, and $\hat{\mu}_j = \Sigma_{t=1}^T \hat{u}_t{}^j/T$ for $j = 3, 4$. The first term of this statistic is in fact a measure of the skewness that reflects departure from a symmetric distribution, while the second term is an excess kurtosis coefficient that measures any excess thickness in the tails of a distribution, relative to that of a normal distribution.

While it is generally the case that applied researchers use formula (2.19) to carry a test of departure from normality. A test of asymmetry can be carried out with the following statistic:

$$\frac{T}{6}\cdot\left[\frac{\hat{\mu}_3{}^2}{(\hat{\sigma}^2)^3}\right] \overset{a}{\sim} \chi^2(1); \tag{2.20a}$$

while a test of excess kurtosis can be carried out as:

$$\frac{T}{24}\cdot\left(\frac{\hat{\mu}_4}{(\hat{\sigma}^2)^2} - 3\right)^2 \overset{a}{\sim} \chi^2(1). \tag{2.20b}$$

---

[25] Harvey (1990) quotes the following remark attributed to the mathematician Poincarré: "...everyone believes in the (Gaussian) laws of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an empirical fact."

In practice, notice that the LM test of normality is very sensitive to outliers amongst the observations. It is therefore advisable to examine histograms of the residuals as they may suggest that the rejection of normality may be accounted by the presence of significant outliers. If it is possible to explain outlying observations by the occurrence of a particular event, dummy variables may be added to the model to capture the behaviour of these particular observations.

2.3.1.4 Joint tests.

The statistical inadequacy tests that we have described above were concerned with testing for one econometric problem at a time. Unfortunately, nothing guarantees that an empirical model cannot be plagued by several inadequacies at once. In some cases, we have proposed how to remedy or to modify a test so that it can be robust to other specific violations of the fundamental conventions underlying the basic general linear model. Another issue concerns how would diagnostic tests of each particular convention be optimally combined under the null hypothesis that there is none of the problems being investigated, i.e. how to perform simply joint tests of inadequacies.

A basic theorem in statistics says that if two independent random variables, say W_1 and W_2, are each distributed according to a χ^2 distribution with n_1 and n_2 degrees of freedom respectively, then $W_1 + W_2 \sim \chi^2(n_1 + n_2)$. Most diagnostic tests that were discussed in this chapter follow a chi-squared distribution. Hence, we could possibly test jointly for evidence of many problems by adding them up, provided that the univariate tests are independent from one another. If not, we need to estimate jointly all the auxiliary regressions associated with a specific violation of a convention, and to carry the joint test over the whole system of equations.

Table 2.3, adapted from Pagan and Hall (1983), reports under which conditions the independence of different diagnostic test statistics prevails so that the individual tests may be added up. Under the heading of specification tests, we consider all the tests discussed as augmenting the conditional mean equation but the serial correlation test. Notice also that the following tests are interpreted on asymptotic grounds as a rule. In many cases, their small sample behaviour is not well known and would require Monte-Carlo experiments to construct the finite sample distributions.

It is therefore advisable to look sometimes at joint departures from the conventions underlying least-squared estimators especially to determine the size of the tests one wants to carry.[26] Furthermore, there seems to be an *order* in which individual tests should be looked at. Specification tests need to be given a priority consideration as the second and higher moment tests need that the conditional mean be correctly modelled. In a similar fashion, the third- and higher moments tests rely upon an adequate statistical representation of the second moments, etc.

Table 2.3- Independence and combination of diagnostic tests.

If $W_i \sim \chi^2(n_i)$ for different i such that W_i are independent, then $\Sigma_i W_i \sim \chi^2(\Sigma_i n_i)$.

Nature of the joint diagnostic tests	Required conditions
Specification test + Serial correlation test of order q	A1. No lagged value of the dependent variable of order less than q (i.e. no y_{t-k} for $k \leq q$) are elements of X, or of the added variables Z in the specification test.
Specification test + Heteroskedasticity test	A2. The distribution of the error term must be symmetric. (Notice that normality of u_t is sufficient but not necessary for symmetry).
Specification test + Normality test	None.
Serial correlation test of order q+ Heteroskedasticity test	A3. Symmetry of the error term or that no lagged value of the dependent variable of order less than q (i.e. no y_{t-k} for $k \leq q$) are elements of X, or of the added variables Z in the heteroskedasticity test. (Notice that normality of u_t is sufficient but not necessary for symmetry).
Serial correlation test + Normality test	None.
Heteroskedasticity test + Normality test	None.
Specification test + Serial correlation test of order q + Normality test	A1.
Specification test + Heteroskedasticity test + Normality test	A2.
Serial correlation test of order q+ Heteroskedasticity test + Normality test	A3.
Specification test + Serial correlation test of order q + Heteroskedasticity test + Normality test	A1. and A2.

---

[26] If one performs J individual tests with respective significance levels of α_j, the level of significance of a joint test of all the single hypotheses is given by $[1 - ((1-\alpha_1)(1-\alpha_2)(1-\alpha_3) \cdots (1-\alpha_J)]$. For instance, two individual hypotheses tested separately at a significance level of 5% amounts to a joint test at the $[1 - 0.95^2] = 9.75\%$ significance level.

2.3.2 Statistical inadequacy tests for structural equations.

We have considered statistical inadequacy tests that one should use when estimating an empirical model with OLS. The importance of testing for statistical adequacy before beginning inference regarding economically interpretable hypotheses remains so even when considering a simultaneous equation model, or a structural equation that could be part of a larger system and that includes some endogenous regressors. Godfrey (1988) reviews many of the available tests for both situations.

In this section, we will concentrate on individual structural equation with some endogenous regressors. Fortunately, the same type of diagnostic tests based on the variable addition principle that were discussed for OLS regressions can be adapted for empirical models estimated by the GIVE procedure.

Let us consider the general linear model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \tag{2.21}$$

where $\mathbf{X}$ is a $k \times T$ matrix but under the assumption that $\mathbf{X}$ and $\mathbf{u}$ are not independent. Let us suppose however that there are $s \geq k$ predetermined variables that may form a an $s \times T$ matrix $\mathbf{W}$, such that these instruments are asymptotically independent from $\mathbf{u}$. We would then use the GIVE procedure to obtain $\hat{\beta}_{GIVE}$.

In this case, specification tests on the conditional mean can be carried out by constructing a Wald (or *F*-test), LM or LR test of H_0: $\gamma = 0$ using either of the following equations that would have been estimated by GIVE:

$$\mathbf{y} = \mathbf{X}\,\beta + \tilde{\tilde{\mathbf{Z}}}\,\gamma + \varepsilon \tag{2.22}$$

or

$$\hat{\mathbf{u}}_{GIVE} = \mathbf{X}\,b + \tilde{\tilde{\mathbf{Z}}}\,\gamma + \mathbf{v} \tag{2.23}$$

where $\hat{\mathbf{u}}_{GIVE} = \mathbf{y} - \mathbf{X}\,\hat{\beta}_{GIVE}$, and $\tilde{\tilde{\mathbf{Z}}}$ is $q \times T$.

If the software at one's disposal does not have a pre-programmed command for Wald, LM, or LR tests with GIVE estimates, a relatively easy way to compute a Wald test of H_0: $\gamma = 0$ is to follow the next steps:

1. First, estimate the original model by GIVE and get the GIVE residuals, i.e. $\hat{\mathbf{u}}_{GIVE}$.
2. Second, regress all columns of X and $\tilde{\mathbf{Z}}$ on W, then construct the predicted values: $\hat{\mathbf{X}}$ and $\hat{\tilde{\mathbf{Z}}}$.
3. Estimate by OLS the auxiliary regression
   $\hat{\mathbf{u}}_{GIVE} = \hat{\mathbf{X}}b + \hat{\tilde{\mathbf{Z}}}\,\gamma + \mathbf{v}$
   and get the R^2 of this regression.
4. To test $H_0\colon \gamma = 0$ with a statistic that asymptotically follows a χ^2-distribution with q of freedom, calculate:
   $Wald = T \cdot R^2/(1-R^2) \sim \chi^2(q)$.

Another possibility to test $H_0\colon \gamma = 0$ is to construct a GIVE analogue to the likelihood ratio (LR) statistics:

1. Estimate the original model by GIVE (which corresponds to the restricted model), get the GIVE residuals, $\hat{\mathbf{u}}_{GIVE}$, and compute the GIVE minimand: $Q^R_{GIVE} = \hat{\mathbf{u}}'_{GIVE}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\hat{\mathbf{u}}_{GIVE}$.
2. Estimate by GIVE the auxiliary regression (which corresponds to the unrestricted model): $\mathbf{y} = \mathbf{X}\beta + \tilde{\mathbf{Z}}\gamma + \varepsilon$. Get the GIVE residuals, $\hat{\varepsilon}_{GIVE}$, and compute the GIVE minimand:
   $Q^U_{GIVE} = \hat{\varepsilon}'_{GIVE}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\hat{\varepsilon}_{GIVE}$.
3. Compute an efficient estimator σ^2_u under the null hypothesis, such as $\hat{\sigma}^2_{GIVE} = T^{-1} \cdot \hat{\mathbf{u}}'_{GIVE}\hat{\mathbf{u}}_{GIVE}$.
4. To test $H_0\colon \gamma = 0$ with a statistic that asymptotically follows a χ^2-distribution with q of freedom, calculate:
   $LR = \left(Q^R_{GIVE} - Q^U_{GIVE}\right)/\hat{\sigma}^2_{GIVE} \sim \chi^2(q)$.

In practice, one is constrained in the dimension q of $\tilde{\mathbf{Z}}$ that can be considered, since for the GIVE to be applicable, one needs the dimension of the matrix of instruments to be greater or equal than the number of regressors in the auxiliary regressions, i.e. $s \geq k+q$. In some cases, it might be possible to augment the set of instruments; in others, it will be necessary to drop some of the regressors in $\tilde{\mathbf{Z}}$. Notice that the asymptotic distribution of the test statistics under the null hypothesis will not be affected by changes in the chosen instruments, for a given number of added up regressors, provided that the instruments are asymptotically independent of the error term. Meanwhile, there may be an effect on the asymptotic local power of the tests as well as differences in small sample performance.

To apply a RESET test in the case of a GIVE regression, one could define $\tilde{\mathbf{Z}} \equiv \left[\hat{y}^2_{GIVE}, \hat{y}^3_{GIVE}, \hat{y}^4_{GIVE}\right]$. Pagan and Hall (1983) suggest also to construct $\tilde{\mathbf{Z}}$ as the powers of the predictions of y from its regression on the predetermined variables of the original equation. Notice that the construction of the proxy variables for the RESET test and the use of GIVE validates the inclusion of the proxy variables amongst the set of instruments in the estimation of the auxiliary regression.

A serial correlation test of order q involves the use of the additional regressors $\tilde{\mathbf{Z}} \equiv [\hat{u}_{GIVE,-1}, \hat{u}_{GIVE,-2}, \ldots, \hat{u}_{GIVE,-q}]$ and, if necessary, he or she can augment the set of instruments by selecting some of the non-redundant pre-determined variables amongst $\mathbf{y}_{-1}$, $\mathbf{X}_{-1}$; ...;$\mathbf{y}_{-q}$, $\mathbf{X}_{-q}$.[27, 28]

To apply a one-step ahead predictive failure test, there are however some complications. If the potential source of instability in the parameter coefficients were restricted to predetermined coefficients, then $\tilde{\mathbf{Z}}$ could be defined as in the test proposed for the OLS regression, except that it would be carried in the GIVE context. The list of instruments would be augmented by adding in $\tilde{\mathbf{Z}}$. Unfortunately, as it is likely, if there is instability in the regression coefficients of the endogenous variables, the usual test is inappropriate as we would need predictions for reduced forms underlying these endogenous variables. Other options are available in some circumstances and are reviewed by Godfrey (1988).[29]

Following Sargan (1964), a general test of misspecification can also be constructed as a test of the overidentifying restrictions if the number of instruments is greater than the number of parameters to be estimated (i.e. $s > k$). That is, under the hull hypothesis that the model is correctly specified and that the s instrumental variables W are valid instruments, then

$$\chi^2_{SM} = Q_{GIVE}(\hat{\beta}_{GIVE})/\hat{\sigma}^2_{GIVE} \sim \chi^2(s-k), \tag{2.24a}$$

---

[27] What is meant by non-redundant is that the extra instruments should not be already used as regressors in the model.

[28] Breusch and Godfrey (1981) have also developed a serial correlation test for structural equations that is based on Sargan (1964) general misspecification test. It requires programming with linear algebra and is written as $\chi^2_{sm}(q) = \hat{u}_{GIVE}{}' Z(Z'HZ)^{-1}Z' \hat{u}_{GIVE} / \hat{\sigma}^2_{GIVE} \sim \chi^2(q)$, where Z is the $T \times q$ matrix of the q lagged values of $\hat{u}_{GIVE}$, with zeros in placed of missing values, and $\mathbf{H} = \mathbf{I}_T - \mathbf{X}(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}' - \hat{\mathbf{X}}(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\mathbf{X}' + \mathbf{X}(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\mathbf{X}'$ in which $\hat{\mathbf{X}} = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}$.

[29] For instance, if the number of observations in each sub-sample is greater than the number of instruments, Lo and Newey (1985) have proposed to test the stability of the coefficient vector with the following statistics: $(\hat{\beta}_{GIVE(I)}-\hat{\beta}_{GIVE(II)})'[\hat{V}_{GIVE(I)} - \hat{V}_{GIVE(II)}](\hat{\beta}_{GIVE(I)} - \hat{\beta}_{GIVE(II)}) \sim \chi^2(k)$, where $\hat{\beta}_{GIVE(j)}$ and $\hat{V}_{GIVE(j)}$ are respectively the GIVE estimate of the coefficients for sample j and the estimator of its asymptotic variance-covariance matrix, for $j = I, II$.

where $Q_{GIVE}(\hat{\beta}_{GIVE}) = \hat{u}'_{GIVE} W(W'W)^{-1} W' \hat{u}_{GIVE}$. The statistic defined in (2.24a) can easily be computed using the R^2 obtained from the regression of $\hat{u}_{GIVE}$ on the matrix of instruments W since:

$$\chi^2_{SM} = T \cdot R^2 \sim \chi^2(s-k). \qquad (2.24b)$$

It is important to notice that in dynamic models with lagged endogenous variables, if there is autocorrelation of the error term, then using values of these lagged endogenous variables in W as if they were exogenous will be a specification error. This may lead the test of overidentifying restrictions to reject the null hypothesis of no misspecification. As emphasized by Hansen (1982), this test is valid for independently and identically distributed disturbances. This suggests that if lagged values of regressors are included in a dynamic model, the test of overidentifying restrictions should be conducted in conjunction with serial correlation test on the residuals. It could be informative as well to proceed with Spencer and Berk's (1981) version of the Hausman test on the potentially endogenous lagged regressors.

To test for heteroskedasticity, it is common practice to apply the same statistical tests as with the OLS equation, except that one tests the significance of the slope coefficients in the OLS regression of $\hat{u}^2_{GIVE}$ on unity and $\tilde{Z}$. This is in fact appropriate provided that the disturbances that would be associated with the unspecified structural equations of each endogenous variable were homoskedastic. If not, one needs to program a different variance-covariance estimator of the slope coefficients as described by Pagan and Hall (1983, p. 196).

Normality of the error term in a structural equation is generally tested with the Bera-Jarque statistics using the GIVE residuals. However, this test is strictly appropriate in this case under the assumption that all endogenous variables follow a joint normal distribution, which may be a strong assumption.[30] Nevertheless, the Bera-Jarque test can be reported as an indicator even if is not exactly correct.

2.3.3 Non-nested tests and encompassing tests

There are instances where one has to contrast two rival models that are non nested within each other. For instance, let us consider the following two alternative models:

---

[30] **Alternative methods would require a test of multivariate normality of the endogenous variables. See Mardia (1980) and Jarque and McKenzie (1983).**

$$M_1: \quad \mathbf{y} = \mathbf{X}_1 \beta + \mathbf{u}_1, \tag{2.25}$$

where $\mathbf{X}_1$ is $T \times k_1$ and $\mathbf{u}_1 \sim N(0, \sigma_1^2 \mathbf{I}_T)$, and

$$M_2: \quad \mathbf{y} = \mathbf{X}_2 \beta + \mathbf{u}_2, \tag{2.26}$$

where $\mathbf{X}_2$ is $T \times k_2$ and $\mathbf{u}_2 \sim N(0, \sigma_2^2 \mathbf{I}_T)$. If $\mathbf{X}_1$ (or $\mathbf{X}_2$) were a subset of linear combinations of the columns in the other matrix, it would be easy to write a general specification that would nest the alternative model as a restricted specification. This would in effect impose restricted values of zero on the coefficients of the relevant variables in the general model. Usual test statistics would then be in order. However, in cases where no model is nested in the other, we have to find another way to compare the relative ability of each specification to explain the variation in y.[31] This will arise for instance, if some columns in $\mathbf{X}_1$ are not included in $\mathbf{X}_2$, and at the same time, some columns in $\mathbf{X}_2$ are not included in $\mathbf{X}_1$. This includes the case for which a variable would appear in level in model *1* and in log-level in model *2*.

Mizon (1984) has argued that a preferred model should be able to represent the characteristics of its rivals, so that one should compare the alternative conditional distributions of y given $\mathbf{X}_1$, and of y given $\mathbf{X}_2$. Two sets of partial indices should be examined: a mean encompassing test and a variance encompassing test.[32] If both sets of tests agree, this may help discarding an unaccceptable model.

The mean encompassing test involves the estimation of the comprehensive model:

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{e}. \tag{2.27}$$

To test M_1 against M_2, we want to compare the unconditional estimate of β_2 with the (probability limit) estimate of β_2 given M_1. This can be done with a Wald- (or *F*-) test of H_0: $\beta_2 = 0$ after estimating equation (2.27). To test M_2 against M_1, we similarly compute a Wald test of H_0: $\beta_1 = 0$.[33] In practical work, it may be possible that the mean encompassing test leads to reject one model against the other. In some instances, it may also happen that this procedure suggests that both models are inadequate on their own, as revealed by the separate rejection $\beta_2 = 0$ and of $\beta_1 = 0$.

---

[31] This section provides only a basic presentation of non-nested hypothesis testing. This wide literature is reviewed in Godfrey (1985), and many tests are briefly presented in Pesaran and Pesaran (1991).

[32] As shown by Mizon and Richard (1986), it is also possible to apply a joint mean and variance (or complete) encompassing test.

[33] Of course, if a subset of variables is thought to belong to both M_1 and M_2, the mean encompassing test can be applied respectively only to the distinct subset of regressors.

Second, the variance encompassing test of M_1 against M_2 involves the comparison of the estimate of σ^2_1, against the (probability limit) estimate of σ^2_1 given M_1. This can be done with Davidson and MacKinnon's (1981) nonnested *J*-test, which is justified on asymtotic grounds, and which is constructed as follows:

1. Estimate model M_2, to get $\hat{\beta}_2 = (X_2'X_2)^{-1} X_2' y$;
2. Estimate $y = X_1 \beta_1 + \lambda (X_2 \hat{\beta}_2) + e_1$, and test H_0: $\lambda = 0$;
3. If H_0 is not rejected, then M_1 is not rejected in favour of M_2. If H_0 is rejected, then M_1 is rejected in favour of M_2.

Similarly, the variance encompassing test of M_2 against M_1 is carried out as follows:

1. Estimate model M_1, to get $\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' y$;
2. Estimate $y = X_2 \beta_2 + \mu (X_1 \hat{\beta}_1) + e_2$, and test H_0: $\mu = 0$;
3. If H_0 is not rejected, then M_2 is not rejected in favour of M_1. If H_0 is rejected, then M_2 is rejected in favour of M_1.

Results from the *J*-test may provide evidence against one model in favour of another (if either λ or μ is different from zero), against both specifications (if both λ and μ are statistically different from zero), or against neither (if both λ and μ are not differrent from zero). In finite sample, as summarized in Godfrey (1985), it seems that the performance of the *J*-test in rejecting a false model is superior to that of the *F*-test, but the *J*-test may reject the true model too often at conventional significance levels. This is why it is recommended to conduct the *J*-test with higher significance levels.

2.3.4 Information criteria for selecting regressors

When dealing with regression equations that may include many regressors in a finite sample context, one has to take account of the trade-off between the number of regressors that may increase the degree-of-goodness of fit and the loss in the number of degrees of freedom that leads to a loss in precision of the coefficients of the included regressors. Such instances are particularly frequent in dealing with dynamic models of time-series, as we will discuss in details in later chapters. Usual *F*-tests can evaluate the nominal statistical significance of some subsets of regressors to guide towards the selection of regressors. Also, Theil (1961) had proposed to select the model with the largest $\bar{R}$, which amounts to prefer the model with the smallest standard-error of the residuals. This criterion should on average pick the correct model, provided that it is amongst the models considered in the analysis. Alternatively, there are instances for which one may instead base his selection of regressors on some information criteria that seek to minimize the mean-squared errors of prediction while penalizing to some extent for too large a set of regressors. Table 2.4 shows several information criteria.[34, 35]

A model with k_j regressors would therefore be selected if it minimized the information criteria under consideration. Notice that the preferred models may not always be the same for each criteria, as they differ in the penalty they adopt for losses in degrees of freedom. In particular, *SC* penalizes heavily for extra regressors. In some instances however, as shown in Ramanathan (1989), some criteria behave similarly across different specifications. If the exclusion of one or more regressors results in a lowering of $\hat{\sigma}^2$, then *GVC* and *RICE* will decrease as well. If the value of *AIC* diminishes when dropping some regressors, SC will

Table 2.4- Various information criteria for selecting regressors, choosing k_j that minimizes the information criterion.[1]

Theil's (1961) $\hat{\sigma}^2$	$\frac{SSE_j}{T-k_j}$	Hannan and Quinn's (1979) *HQ*	$\frac{SSE_j}{T}(\log T)^{2k_j/T}$
Akaike's (1973) *AIC*	$\frac{SSE_j}{T}\exp\left(\frac{2k_j}{T}\right)$	Craven and Wahba's (1979) Generalized Cross Validation *GVC*	$\frac{SSE_j}{T}\left[1-\frac{k_j}{T}\right]^{-2}$
Mallows' (1973) C_p	$SSE_j + 2\ k_j\ \hat{\sigma}^2_m$	Amemiya's (1980) *PC*	$SSE_j\left(\frac{T+k_j}{T-k_j}\right)$
Akaike's (1970) Finite Prediction Error *FPE*	$\frac{SSE_j}{T}\left(\frac{T+k_j}{T-k_j}\right)$	Shibata's (1981) *SHIBATA*	$\frac{SSE_j}{T}\left(\frac{T+2k_j}{T}\right)$
Schwarz's (1978) *SC*	$\frac{SSE_j}{T}(T^{k/T})$	Rice's (1984) *RICE*	$\frac{SSE_j}{T}\left[1-\frac{2k_j}{T}\right]^{-1}$

[1] SSE_j: residual sum of squares for the j^{th} model; $\sigma^2_m = SSE_m / (T-k_m)$, i.e. for the most general model, with all the k regressors.

---

[34] For excellent surveys of these criteria, see Engle and Brown (1985), Ramanathan (1989) and Maddala (1992).

[35] Some of the criteria shown in this table are sometimes expressed equivalently in terms of the value of the maximized log-likelihood function. For instance, $AIC = (-2 \log L + 2\ k)/T$.

decrease if $T \geq 8$, and *HQ* will decrease if $T \geq 16$. Finally, if excluding regressors rise the value of *RICE, SHIBATA* also increases.

In constrast, with traditional statistical tests at a 5% significance level, information criteria generally amounts to tests conducted with a higher significance level. On the other hand, these criteria are also generally lowering the level of significance with the sample size *T*.[34]

While such criteria may be helpful in discarding less useful specifications, an adopted empirical model should still be submitted to the statistical tests for inadequacy that were discussed earlier in this chapter.

2.4 An Econometric Approach to Model Selection: Hendry's Methodology

> "The three golden rules of econometrics are test, test, and test."
>
> David Hendry (1980)

Traditionally, applied economists have often favoured the estimation of highly parsimonious and specific models at the outset, that were evaluated in their capacity to fit the data in a way compatible with some economic theory. If the reasearcher did not feel comfortable with some aspects of the specification, e.g. because of too small a goodness-of-fit or because of a wrong sign on some estimated coefficient, then the search involved adding additional regressors. This *specific to general* approach is inappropriate for many reasons. Tests conducted first on a more specific model maintain zero coefficient restrictions on additional regressors without having tested them. If these maintained restrictions are false, it is likely that the inference conducted on the coefficients of the included regressors will be biased. Also, because the sequence of hypothesis testing is conducted in some unstructured manner, the significance levels of this testing procedure are unknown.

By contrast, Hendry (1979) has advocated a *general to specific* approach, from a possibly overparameterized model. A search for an appropriate model involves therefore testing down the general model to simplify the model based on the evidence gathered from the data.

---

[34] This can be particularly appropriate when one is working with a high value of *T*, since it is has been shown that a *too-large sample size problem* can arise at times. That is, with a large enough *T*, almost any coefficient estimates can be found to be statistically significant even if the magnitude of the coefficient is small. See Kennedy (1992), p. 65.

Such a sequence of testing has the advantage that one knows the significance level of the j^{th} test that is performed, so that the overall probability of the Type I error is controlled.[37] In fact, as one applies further tests, down his search path for a statistically adequate model, it is advisable to lower the significance level of the tests below 5%. Or, equivalently, one should perform joint tests at some specified significance level. Futhermore, the application of misspecification tests should also be part of the model selection procedure.

Hendry's approach to applied econometrics rests on the idea that good empirical work requires that one's model must reasonably satisfy a wide list of criteria that makes a chosen model congruent with the evidence. Table 2.5 provides a checklist of criteria and associated methods that should guide a careful applied researcher. There are mainly five categories of criteria. Consistency implies that a model should have logically possible signs and magnitudes of parameters, and of predictions. It should also be consistent with existing long-run equilibria between the variables. The estimated model should exhibit economic and statistical significance of its parameters. Various indices of inadequacy should suggest that the model provides an adequate representation of the data. It should be encompassing characteristics of rival models, so that the latter contain no information that would be useful to improve the preferred model. Finally, the model should not display too much sensitivity to the sample size, the variable menu or to other equations of the system (when applicable).[38] To perform a sensitivity assessment of the model with respect to the sample size, it is highly recommended that a researcher keeps aside a subsample of observations that will not be used in the specification search/estimation stage. Then, the model can be evaluated regarding its capacity to predict out-of-sample observations.

---

[37] See footnote no 26 in this chapter.

[38] Sensitivity to the menu of regressors, or fragility analysis, has been recommended by Leamer (1983). An interesting application is that of McAleer, Pagan and Volker (1985).

Table 2.5 - Model Selection Criteria[1]

Category	Property	Examples/Methods
CONSISTENCY	(a) Inadmissibility	Signs, magnitude of parameters and predictions.
	(b) Poor operating characteristics	No flow equilibria; unacceptable (implicit) stock equilibrium. Inconsistent with long-run equilibrium.
SIGNIFICANCE	(a) Economic	Quantitative impact of unacceptable magnitude.
	(b) Statistical	
	(i) Nominal significance levels	*F*-tests, *t*-tests, Wald tests, etc.
	(ii) Optimized significance levels	Various information criteria: e.g. AIC, SC, GVC, etc.
INDICES OF INADEQUACY	(a) For the conditional mean	RESET, *LM*-test for serial correlation, PSW differencing test, Hausman test
	(b) For the conditional variance	Various heteroskedasticity tests: e.g. White, Ramsey, ARCH
	(c) For normality	Bera-Jarque test
ENCOMPASSING	(a) The mean	*F*-test
	(b) The variance	*J*-test
SENSITIVITY	(a) To sample size	Chow- test, One step-ahead prediction error test.
	(b) To variable menu	Extreme bound analysis (See Leamer (1983)).
	(c) To equations of model (when applicable)	Simulation

---

[1] Adapted from Adrian Pagan's course Economics 517 Model Selection and Evaluation, taught at the University of Rochester in the Fall 1987.

2.5 Summary

The object of this chapter was threefold. First, it aimed at reviewing intuitively fundamental econometric concepts, regarding the basic econometric conventions and properties of commonly used estimators, as well as the consequences occurring from departures from these conventions. Second, it developped a set of tools around the general principle of variable addition, which allowed to cast several very important test statistics to assess the statistical adequacy of empirical models into a unified framework. Third, these tests were showed to play an active role within a sound and general approach to empirical work, namely Hendry's methodology.

Essentially, good applied work that seeks to uncover economic relationships in the data must involve a systematic and careful use of econometric techniques that treat the data with respect. This requires a thorough knowledge of the strengths and weaknesses of these techniques. Furthermore, it is strongly recommended to begin one's quest with a general specification that will be simplified on the basis of the evidence revealed by the data. This evidence should be considered revealing only insofar as a model is an adequate statistical representation according to some statistical criterion. This is why it is essential to report results from a battery of misspecifiation tests on a selected model. These tests should consider in turn, or jointly, possible inadequacies in the conditional mean, in the conditional variance, and in normality. It is only at this stage that one should test interesting economic hypotheses and draw some economic conclusions about the data. The sound application of this econometric methodology for model building should provide more thrust in the econometric results one obtains, and less unpleasant surprises in the quality of one's understanding of economic data and their forecasts.

3. THE CHARACTERIZATION OF UNIVARIATE TIME SERIES.

3.1 Introduction.

As macroeconomists seek to explain the evolution and comovements of aggregate variables over time, the data we mostly work with involve sequences of observations ordered in time, called time series. The observed data can be thought of as resulting from an underlying stochastic process itself characterized by some law of probability. One goal is thus to know the data at hands before embarking in a search of specific relationships, between different macroeconomic variables. Also, we might be interested for its own sake in identifying the nature of the data, and particularly in knowing whether the data are stationary or not. Finally, for forecasting purposes alone, it may be satisfactory enough to construct and to estimate an empirical model describing the evolution of a variable with its past values and with past and current values of its innovations.

This chapter reviews the use of helpful tools necessary to characterize and to represent the univariate behaviour of time series through identification, estimation and diagnostic checking. Special attention is devoted to a particular form of non-stationarity in macroeconomic time series, namely the existence of unit roots. This form of non-stationarity and the developments of several unit root tests have stimulated an extensive literature in the 1980's. It is essential for a modern applied macro-econometrician to know about these tests and to be able to understand the extent and in which circumstances unit roots matter.

3.2 Identifying, Estimating and Diagnostic Testing.

3.2.1 Introduction

The characterization of univariate time series by some linear stochastic representation is essentially an iterating procedure that can be defined in three stages. At first, it requires to discard most of the unhelpful and inaccurate representations to restrict initially one's study to one or just a few potentially satisfactory models. This is called identification. The second stage consists in the estimation of retained models. Finally, diagnostic tests are applied to check whether the modeller's choice is not invalidated. If the model is rejected, we must go back to the identification stage to look for alternative set-ups that would be satisfactory.

3.2.2 *Definitions and theoretical properties of stationary stochastic processes*

Let us consider a sample of T observations $\{y_1, y_2, ..., y_T\}$ for a stochastic variable Y_t. That is, Y_t is determined at least partially by a random process. This realization of the process can be resulting from a joint probability distribution $p(y_1, y_2, ..., y_T)$, whose theoretical moments for each period t include a mean $E[Y_t] = \mu_t$, a variance $\gamma_{tt} \equiv E[(Y_t - \mu_t)^2] = \sigma_t^2$, and autocovariances $\gamma_{ts} \equiv E[(Y_t - \mu_t)(Y_s - \mu_s)]$. If we were dealing with some reproducible drawings, we could collect an ensemble of realizations, i.e. many samples from 1 to T, and compute some sample counterparts to the theoretical moments mentioned above. For each observation t, we would compute estimates of the mean, of the variance and of the autocovariances.

However, history has unfolded only once and has therefore provided the econometricians with only one realization of a time series. This is why we need to rely on some assumptions to uncover the statistical structure of the underlying stochastic process, using a single sample of data collected over some period. Namely, it is commonly assumed that the joint probability distribution of the process is covariance stationary, that is:[1]

1. The mean of the process is independent of time as it is constant: $E[Y_t] = \mu, \ \forall t$;
2. The variance of the process is independent of time as it is constant: $\gamma_{tt} \equiv \gamma_0 = \sigma^2, \forall t$;
3. The autocovariances of the process are independent of time itself and depend only on the time interval between two terms: $\gamma_{ts} \equiv E[(Y_t - \mu)(Y_s - \mu)] = \gamma_{|s-t|} = \gamma_\tau$, where $\tau = s\text{-}t$, $\forall s \neq t$. In particular, $\gamma_\tau = \gamma_{-\tau}$.

Covariance stationary is thus very useful. It allows an applied researcher to construct a class of time-series models from studying the behaviour of a sequence of observed data that may not be independent from one another even though the linear properties of the data are independent of time itself.[2]

---

[1] In fact, if the joint probability distribution function of the process is normal, then the mean and the autocovariances are sufficient moments to characterize completely the distribution. Hence, covariance stationarity implies strong stationarity of the distribution, i.e. all higher moments are also independent of time itself. For non-normal processes however, weak stationarity does not ensure strong stationarity.

[2] Technically, the stochastic process under scrutiny must also be ergodic, i.e. any observed values of a process that are sufficiently far in time from one another must be almost uncorrelated. This requirement concerns the memory of the process, stating that is must not be too long. Ergodicity is essential so that in computing, say the mean of a process each additional observation brings some new additional information that is useful to estimate it. For instance, let $Y_T = T^{-1} \Sigma^T_{t=1} y_t$, $E[Y_T] = \mu$ and $\lim_{T \to \infty} Var[Y_T] = 0$ for all T. Notice that it is not possible to test for ergodicity. However, a necessary albeit non-sufficient condition for ergodicity is that $\lim_{\tau \to \infty} \gamma_\tau = 0$ quickly enough as τ increases.

Because the values of the autocovariances are not independent of the measurement units of the variable, it is common practice, as well as convenient, to normalize the autocovariances by the variance of the process which gives the autocorrelation function:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} \qquad \text{for } \tau = 0, 1, 2, \ldots . \qquad (3.1)$$

Covariance stationarity implies that $\rho_\tau = \rho_{-\tau}$ and there is a one-to-one correspondence between a stationary normal process and a specific autocorrelation function. Notice also that $\rho_0=1$ and $-1 \leq \rho_\tau \leq 1$. There may be however several non-normal processes that share an identical autocorrelation function. Nonetheless, the autocorrelation function is an acceptable and useful way to characterize a stochastic process.

In the class of linear covariance stationary process, a general representation for a variable y_t in terms of its past values, and of past and current values of its innovations is given by:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \quad (3.2)$$

where ε_t is a random variable, called the innovation, which is identically and independently distributed with a zero mean and a constant variance, σ_ε^2. Because of its constant mean and because $E(\varepsilon_t \varepsilon_{t+s}) = 0$, for all $s \neq 0$, the innovation is called a zero-mean white noise process. Since y_t is autoregressively written with p lagged values of itself and with a weighted moving average of the lagged values of its innovations up to order q, the process represented by equation (3.2) is called an *ARMA(p, q)*. In particular, if no lagged values of the innovations contribute directly in determining the current value of y_t, then $y_t \sim ARMA(p, 0) \equiv AR(p)$. If no lagged values of y_t contribute in defining the current value of y_t, then $y_t \sim ARMA(0, q) \equiv MA(q)$. Finally, if $y_t = \delta + \varepsilon_t$, then y_t is identically and independently distributed, such a process is an *ARMA(0, 0)*, i.e. a white noise process.

A notation often encountered in time series papers involves the use of the so called lag operator L. This operator has the properties that it can be treated as any algebraic quantity and that if it is raised to the power j, it shifts a variable backward in time by j-periods. Hence, $y_{t-j} \equiv L^j y_t$. In particular, if the lag operator is applied on a constant, say c, it is non operand, as if it were equal to 1, i.e. $L^j c = c$. Using this notation, we can rewrite equation (3.2) as:

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)\, y_t = \delta + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t. \quad (3.3)$$

This can also be represented as:

$$\phi(L) y_t = \delta + \theta(L)\varepsilon_t, \quad (3.4)$$

where the polynomials in L, $\phi(L)$ of order p, and $\theta(L)$ of order q are respectively associated with the *AR* and the *MA* part of the process. The mean of the process can be reckoned as:

$$\phi(L) E(y_t) = \delta + \theta(L)\, E(\varepsilon_t) \quad \Rightarrow \quad \mu = \frac{\delta}{\phi(L)} = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}, \quad (3.5)$$

since $E(y_t) = \mu$ and $E(\varepsilon_t) = 0$ for all t. To simplify the notation in our theoretical exposition, we will suppose below that a series y_t is already defined in terms of the deviation of the original series from its mean. In practice, the inclusion of an intercept will accomplish the same thing.

A stochastic process *ARMA*(p, q), such as

$$\phi(L) y_t = \theta(L)\varepsilon_t \quad (3.6)$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$, can be rewritten in two ways.

First, writing y_t solely in terms of its past innovations yields its MA(∞) representation given by:

$$y_t = \frac{\theta(L)}{\phi(L)}\varepsilon_t = \psi(L)\varepsilon_t = (1 + \psi_1 L + \psi_2 L^2 + \cdots)\varepsilon_t. \quad (3.7)$$

Furthermore, if this *MA*(∞) representation does not explode, we say that the time series y_t is generated by a stationary process. A necessary condition for an *ARMA*(p, q) to be stationary is that $\Sigma_{i=1}^{p} \phi_i < 1$. A necessary and sufficient condition for stationarity is that all the roots of the lag polynomials $\phi(L)$ have a modulus greater than 1.[3] Notice that the conditions of stationarity depend only upon the finite order polynomial associated with the *AR* part.[4] We will see later that if the required conditions for stationarity are not satisfied, the variance of the process is unbounded.

---

[3] For the $\phi(L)$ polynomial of order p, the solutions of $\phi(L) = (1 - \phi_1 L - \ldots - \phi_p L^p) = (1 - a_1)(1 - a_2) \cdots (1 - a_p) = 0$ give p roots denoted as $1/a_1$, $1/a_2$, ... $1/a_p$. If these roots are real, stationarity requires that their absolute values be all greater than one. If some of these roots are complex, say for instance of the form $A + B\,i$, where $i = (-1)^{0.5}$, stationarity requires that $(A^2 + B^2)^{0.5} > 1$.

[4] Hence, any finite pure MA(q) process will be trivially stationary.

Second, writing the current innovation ε_t in terms of the current and past values of y_t expresses the process as an $AR(\infty)$ representation given by:

$$\varepsilon_t = \frac{\phi(L)}{\theta(L)}\, y_t = \pi(L)\, y_t = (1 + \pi_1 L + \pi_2 L^2 + \cdots) y_t. \tag{3.8}$$

If this $AR(\infty)$ representation does not explode, we say that the variable y_t is generated by a invertible process.[5] A necessary and sufficient condition for invertibility is that all the roots of the lag polynomials $\theta(L)$ have a modulus greater than 1. As we do not observe the true current or past values of the innovations, invertibility is required at the estimation stage so that we can express ε_t in terms of the observed values of y_t.

An often referred basic theorem in time-series is the Wold's theorem that states that any stationary process can be thought of as being uniquely represented as the sum of a purely self-deterministic component and an *MA* component of possibly infinite order. Because of this theorem, a linear representation of a time series exists for all linear or nonlinear time-series.[6] Hence, provided that the series is stationary, any $MA(\infty)$ process can be approximated by some $ARMA(p, q)$ models with a fairly small number of parameters, i.e. so that $p+q$ will not be large. In some instances, pure *AR* or pure *MA* processes may even suffice.

Another tool that is employed in identifying the nature of a stochastic process is the partial autocorrelation function. The partial autocorrelation coefficients $\phi_{\tau\tau}$ of a time series y_t are defined as the sequence of partial correlations between y_t and $y_{t-\tau}$ for $\tau = 1, 2, 3, \ldots,$ conditional upon the given correlation of y_t with $y_{t-1}, \ldots, y_{t-\tau+1}$. That is, approximating the underlying data generating function of y_t by a sequence of AR processes of increasing order $K = 1, 2, 3, \ldots,$ the partial autocorrelations are given by the last coefficient in each of the following autoregressions:

$$\begin{aligned} y_t &= \phi_{11} y_{t-1} + \varepsilon_{1t} \\ y_t &= \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + \varepsilon_{2t} \\ y_t &= \phi_{31} y_{t-1} + \phi_{32} y_{t-2} + \phi_{33} y_{t-3} + \varepsilon_{3t} \end{aligned} \tag{3.9}$$

---

[5] A pure finite AR(p) process will always be invertible.

[6] This comforts most applied econometricians in working with linear models. While this is clearly practical, nonlinear models might well be useful in some circumstances. An example of a nonlinear model is the bilinear model expressed as $y_t = \phi_1 y_{t-1} + \varepsilon_t + \lambda_1 y_{t-1} \varepsilon_{t-1}$. Interested readers are referred to Priestley (1988).

$$y_t = \phi_{K1} y_{t-1} + \phi_{K2} y_{t-2} + \phi_{K3} y_{t-3} + \cdots + \phi_{KK} y_{t-K} + \varepsilon_{Kt}$$

The partial autocorrelation coefficients for $K = 1, 2, ..., \tau$ are given by $\{\phi_{11}, \phi_{22}, \phi_{33}, ..., \phi_{\tau\tau}, ...\}$.[7] Hence, the partial autocorrelation coefficient $\phi_{\tau\tau}$ measures the excess correlation between y_t and $y_{t-\tau}$ not accounted for by the correlations between y_t and its lagged values up to τ-1. Since any ARMA(p, q) can be represented as an AR(∞), the partial autocorrelation coefficients can be theoretically reckoned for any order. However, if the true generating process of a particular time series is given by an AR(p), then it is easy to understand that $\phi_{pp} = \phi_p$, and $\phi_{\tau\tau} = 0$ for all $\tau > p$. Also, notice that by definition, the autocorrelation coefficient of order one is exactly equal to the partial autocorrelation of order one.

The nature of an underlying stochastic process will affect in some particular manner the theoretical characteristics of both the ordinary (ACF) and the partial (PACF) autocorrelation functions. The following set of rules should be considered to identify an ARMA(p, q) process, according to the theoretical behaviour of the ordinary autocorrelation and the partial autocorrelation coefficients:[8]

1. If $y_t \sim$ AR(p), i.e. $q = 0$, then its autocorrelation function declines geometrically and its partial autocorrelation function "cuts off" after some point $\tau = p$;
2. If $y_t \sim$ MA(q), i.e. $p = 0$, then its autocorrelation function "cuts off" after some point τ=q and its partial autocorrelation function dies off geometrically;
3. If $y_t \sim$ ARMA(p, q), then the first q ordinary autocorrelations are a function of both the AR and the MA parameters, while they start declining after some point q, and its partial autocorrelation dies off in some steady fashion.

In particular, the theoretical values of the ordinary and partial autocorrelation coefficients for $\tau > 0$ are all zero for a white noise process. Table 3.1 below shows the theoretical autocorrelation function and the theoretical partial autocorrelation function for different stochastic processes.

---

[7] As shown in most econometric textbooks treating of time-series, each partial autocorrelation $\phi_{\tau\tau}$ can be found as the solution of the simultaneous equation system given by the so called Yule-Walker equations: $\rho_K = \Sigma_{j=1} \phi_{\tau j} \rho_{K-j}$ for K=1, 2, ..., τ. For instance, see Granger and Newbold (1986), and Pindyck and Rubinfeld (1992).

[8] As it is beyond the scope of this guide to derive formally the autocorrelation function and the partial autocorrelation function, the interested reader is referred to time-series textbooks such as Granger and Newbold (1986) and Pindyck and Rubinfeld (1992) or Paquet (1993) for such derivations.

As it is often the case that practical researches involve the use of combinations of time series, the following theorem and its special cases are helpful in understanding the effect of aggregation on the nature of time series. Let us consider two stationary time series, say y_t and x_t, whose underlying stochastic processes are described respectively by an $ARMA(p_1, q_1)$ and an $ARMA(p_2, q_2)$. Then, the time series z_t is defined as the sum of the two original processes

$$z_t \sim ARMA(p_3, q_3) \text{ where } p_3 \leq p_1 + p_2 \text{ and } q_3 \leq max[p_1 + q_2, p_2 + q_1].$$

The definitions of p_3 and q_3 involve inequalities because there are instances where both the polynomials representing the AR and the MA parts of the process share common roots. For instance, let us suppose that z_t has the following process that appears to be an $ARMA(2,1)$:

$$(1 - 0.2L - 0.08 L^2)\, z_t = (1 + 0.2L)\, \varepsilon_t \,. \tag{3.10a}$$

This can be simplified to

$$(1 - 0.4L)\,(1 + 0.2L)\, z_t = (1 + 0.2\, L)\, \varepsilon_t, \tag{3.10b}$$

which implies that

$$(1 - 0.4L) z_t = \varepsilon_t, \tag{3.10c}$$

i.e. z_t follows in fact an $AR(1)$ process. Hence, the original representation of the process in equation (3.10a) was overparameterized.

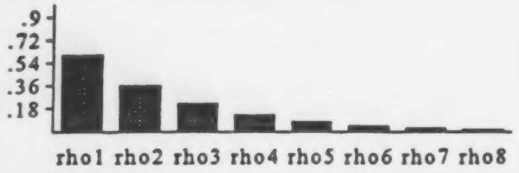
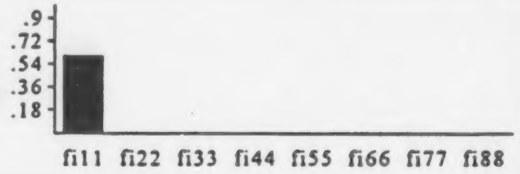
The theorem on aggregation implies the particular results listed below:

- $AR(p) + white\ noise = ARMA(p, p)$
- $AR(p_1) + AR(p_2) = ARMA(p_1 + p_2, max[p_1, p_2])$
- $MA(q_1) + MA(q_2) = ARMA(0, max[q_1, q_2])$
- $ARMA(p, q) + white\ noise = ARMA(p, max[p, q])$
- $AR(p) + MA(q) = ARMA(p, p+q)$.

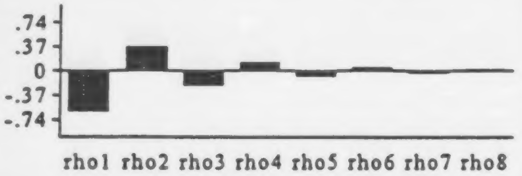
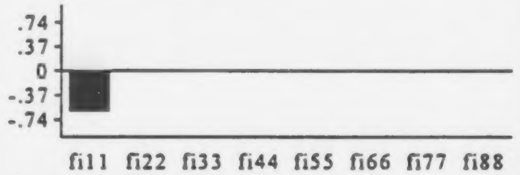
Since we work with data that generally suffer from measurement errors, it is therefore not surprising that actual time series often follow a mixed $ARMA$ process.

Table 3.1. - Theoretical autocorrelation function and partial autocorrelation function for different processes

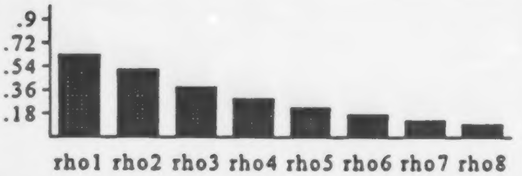
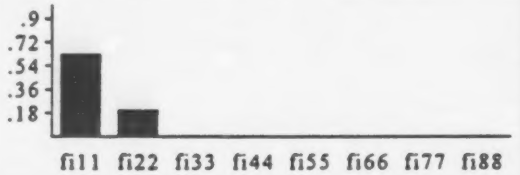
AR(1): $y_t = .6\,y_{t-1} + \varepsilon_t$, roots of $\phi(L)$: |1.67| > 1 => stationary; pure AR => invertible.

ACF	PACF
	

AR(1):: $y_t = -.6\,y_{t-1} + \varepsilon_t$, roots of $\phi(L)$: |-1.67| > 1 => stationary; pure AR => invertible.

ACF	PACF
	

AR(2):: $y_t = .5\,y_{t-1} + .2\,y_{t-2} + \varepsilon_t$, roots of $\phi(L)$: |-3.81| > 1 and |1.31| >1 => stationary; pure AR => invertible.

ACF	PACF
	

AR(2):: $y_t = -.5\,y_{t-1} - .2\,y_{t-2} + \varepsilon_t$, roots of $\phi(L)$: ||-1.25 + 1.85 *i* ||= 2.24 > 1 and ||-1.25 - 1.85 *i* ||=2.24 >1 => stationary; pure AR => invertible.

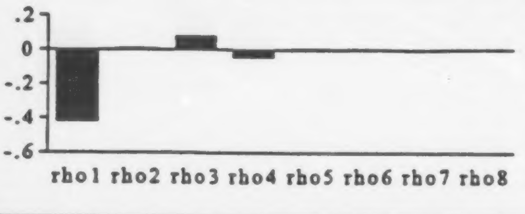
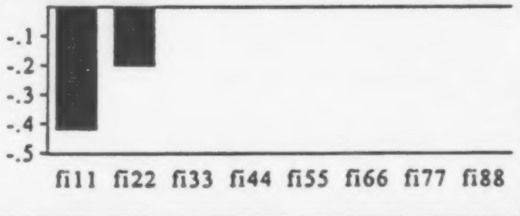
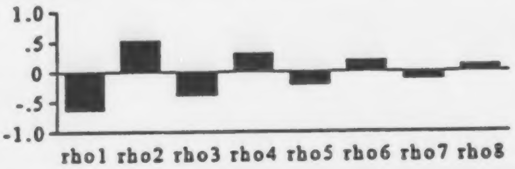
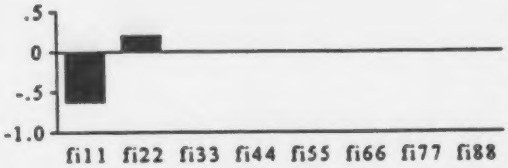
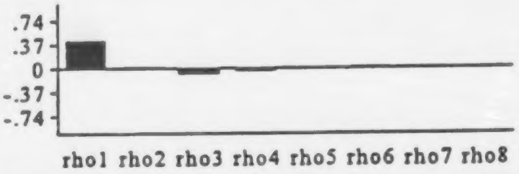
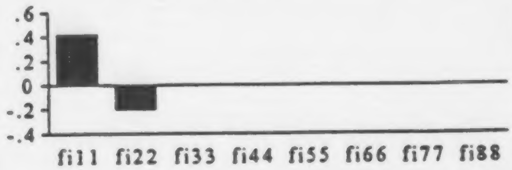
ACF	PACF
	

Table 3.1. - Theoretical autocorrelation function and partial autocorrelation function for different processes

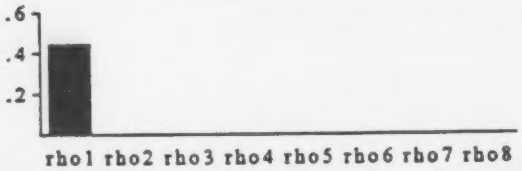
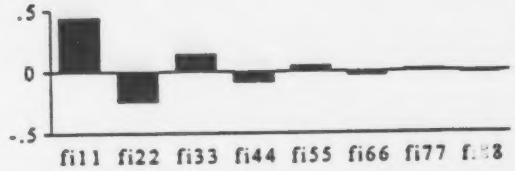
AR(2):: $y_t = -.5\,y_{t-1} + .2\,y_{t-2} + \varepsilon_t$, roots of $\phi(L)$: $\|-1.31\| > 1$ and $\|3.81\| > 1$ => stationary; pure AR => invertible.

ACF	PACF
	

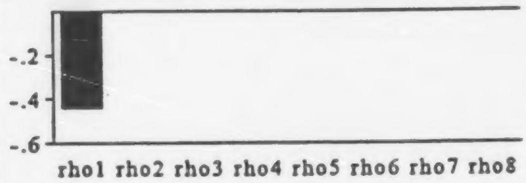
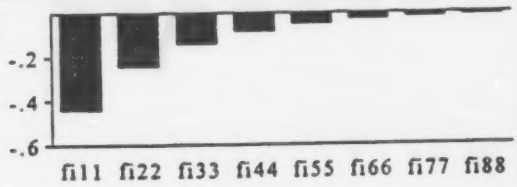
AR(2):: $y_t = .5\,y_{t-1} - .2\,y_{t-2} + \varepsilon_t$, roots of $\phi(L)$: $\|1.25 + 1.85\,i\| = 2.24 > 1$ and $\|1.25 - 1.85\,i\| = 2.24 > 1$ => stationary; pure AR => invertible.

ACF	PACF
	

MA(1):: $y_t = \varepsilon_t + .6\,\varepsilon_{t-1}$; pure MA => stationary; roots of $\theta(L)$: $|-1.67| > 1$ => invertible

ACF	PACF
	

MA(1):: $y_t = \varepsilon_t - .6\,\varepsilon_{t-1}$; pure MA => stationary; roots of $\theta(L)$: $|1.67| > 1$ => invertible

ACF	PACF
	

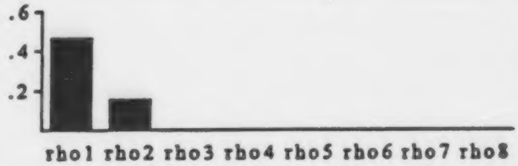
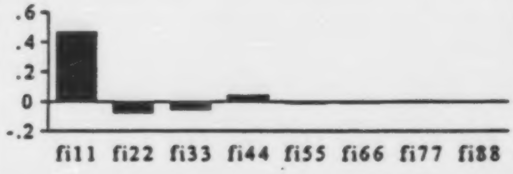
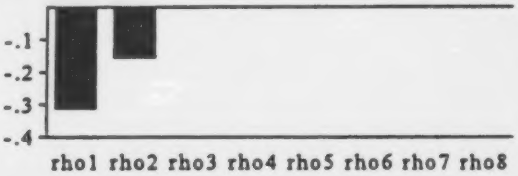
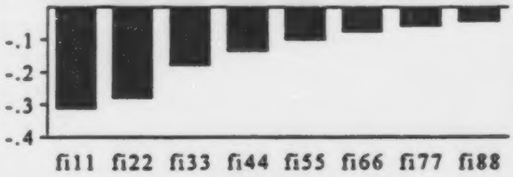
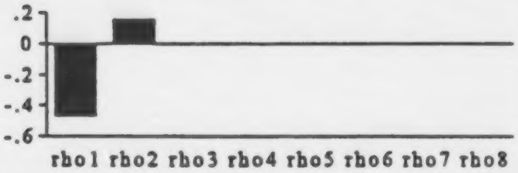
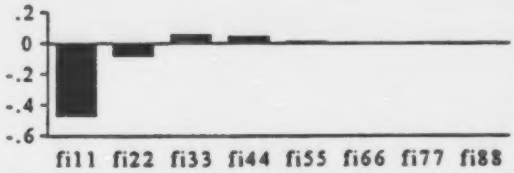
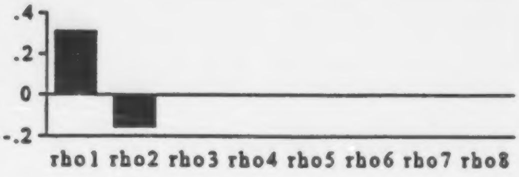
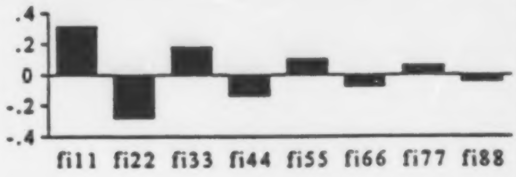
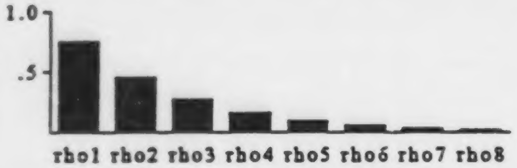
<table>
<tr><th colspan="2">Table 3.1. - Theoretical autocorrelation function and partial autocorrelation function for different processes</th></tr>
<tr><td colspan="2">MA(2):: $y_t = .\varepsilon_t + .5\,\varepsilon_{t-1} + .2\,\varepsilon_{t-2}$; pure MA => stationary;
roots of $\theta(L)$: $\|-1.25 + 1.85\,i\| = 2.24 > 1$ and $\|-1.25 - 1.85\,i\| = 2.24 > 1$ => invertible.</td></tr>
<tr><td>ACF</td><td>PACF</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td colspan="2">MA(2):: $y_t = .\varepsilon_t - .5\,\varepsilon_{t-1} - .2\,\varepsilon_{t-2}$; pure MA => stationary;
roots of $\theta(L)$: $|-3.81| > 1$ and $|1.31| > 1$ => invertible.</td></tr>
<tr><td>ACF</td><td>PACF</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td colspan="2">MA(2):: $y_t = .\varepsilon_t - .5\,\varepsilon_{t-1} + .2\,\varepsilon_{t-2}$; pure MA => stationary;
roots of $\theta(L)$: $\|1.25 - 1.85\,i\| = 2.24 > 1$ and $\|1.25 + 1.85\,i\| = 2.24 > 1$ => invertible.</td></tr>
<tr><td>ACF</td><td>PACF</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td colspan="2">MA(2):: $y_t = .\varepsilon_t + .5\,\varepsilon_{t-1} - .2\,\varepsilon_{t-2}$; pure MA => stationary;
roots of $\theta(L)$: $|-1.31| > 1$ and $|3.81| > 1$ => invertible.</td></tr>
<tr><td>ACF</td><td>PACF</td></tr>
<tr><td> </td><td> </td></tr>
</table>

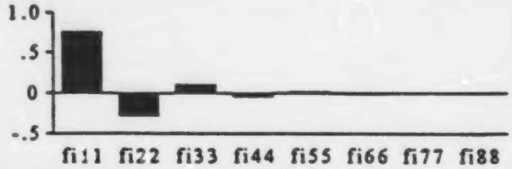
Table 3.1. - Theoretical autocorrelation function and partial autocorrelation function for different processes

ARMA(1, 1):: $y_t = .6 y_{t-1} + \varepsilon_t + .4 \varepsilon_{t-1}$; roots of $\phi(L)$: $|1.67| > 1$ => stationary;
roots of $\theta(L)$: $|-2. 5| > 1$ => invertible.

ACF

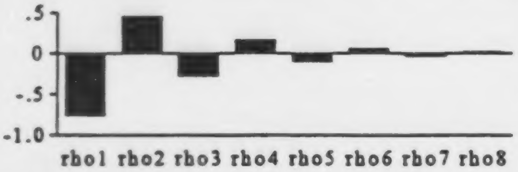


PACF

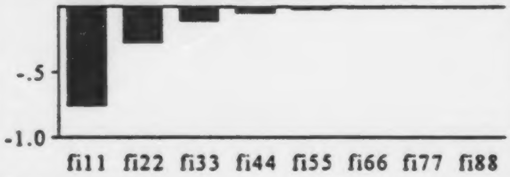


ARMA(1, 1):: $y_t = -.6 y_{t-1} + \varepsilon_t - .4 \varepsilon_{t-1}$; roots of $\phi(L)$: $|-1.67| > 1$ => stationary;
roots of $\theta(L)$: $|2. 5| > 1$ => invertible.

ACF

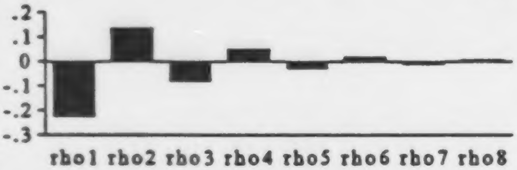


PACF

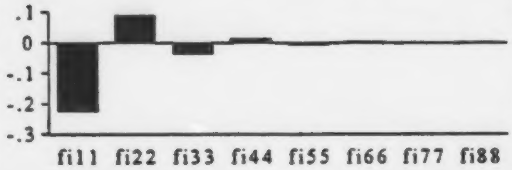


ARMA(1, 1):: $y_t = -.6 y_{t-1} + \varepsilon_t + .4 \varepsilon_{t-1}$; roots of $\phi(L)$: $|-1.67| > 1$ => stationary;
roots of $\theta(L)$: $|-2. 5| > 1$ => invertible.

ACF

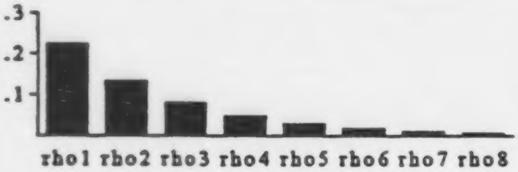


PACF

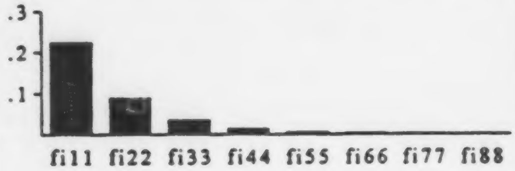


ARMA(1, 1):: $y_t = .6 y_{t-1} + \varepsilon_t - .4 \varepsilon_{t-1}$; roots of $\phi(L)$: $|1.67| > 1$ => stationary;
roots of $\theta(L)$: $|2. 5| > 1$ => invertible.

ACF



PACF



3.2.3 Definitions and theoretical properties of some nonstationary stochastic processes

We have reviewed the characteristics of theoretical ordinary and partial autocorrelation functions of stationary time series. In practice, however, it is often the case that the data we study are nonstationary as they often exhibit trends. For a series y_t, one particular type of nonstationarity is such that if the series is differenced enough times, say d times, the new series is stationary. In this case, it is said that y_t is integrated of order d (noted $I(d)$) or that it is differenced stationary of order d. In terms of the notation that we have defined previously, a time series y_t could be said to follow an $ARIMA(p, d, q)$, i.e.

$$\phi(L)\ (1-L)^d y_t = \theta(L)\ \varepsilon_t, \tag{3.11}$$

where $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$. Hence, the series $w_t \equiv (1-L)^d y_t$ is stationary, even if y_t is nonstationary. In fact, $w_t \sim ARMA(p, q)$.

To get some intuition for this form of nonstationarity inherent in some time series, let us consider the example of $y_t \sim ARIMA(0, 1, 0)$, i.e.

$$y_t = y_{t-1} + \varepsilon_t \qquad \Leftrightarrow \qquad w_t \equiv (1-L)y_t = \varepsilon, \tag{3.12}$$

where $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$.[9] This process can be rewritten as:[10]

$$y_t = \frac{1}{(1-L)}\varepsilon_t = y_0 + \sum_{j=0}^{t-1} L^j \varepsilon_t = y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j} = y_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1 . \tag{3.13}$$

Hence, the current level of y_t is function of its initial level and of all current and past innovations. An innovation that occurred no matter how far in the past still contributes to the time t value of the variable. Another way to illustrate this idea is by computing the variance of the level of the process:

$$Var(y_t) = \gamma_0 = E[(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots \varepsilon_1)^2] = t \cdot \sigma_\varepsilon^2 . \tag{3.14}$$

Hence, as $t \to \infty$, $Var(y_t) \to \infty$, i.e. the variance of y_t explodes.

Notice also that the autocovariances of this process at order $1, 2, \ldots, \tau, \ldots$ are given by:

$$\gamma_1 = (t-1)\sigma_\varepsilon^2; \quad \gamma_2 = (t-2)\sigma_\varepsilon^2; \quad \ldots; \quad \gamma_\tau = (t-\tau)\sigma_\varepsilon^2; \ldots \tag{3.15}$$

---

[9] The following argument would remain valid even if $(1-L)\,y_t$ were an $ARMA(p,q)$.

[10] The algebra in this equation involves the use of the solution of a geometric series, i.e. $S = 1 + z + z^2 + z^3 + \ldots = 1/(1-z)$.

so that the theoretical autocorrelation function of such a process is given by:

$$\rho_0 = 1; \quad \rho_1 = \frac{t-1}{t}; \quad \rho_2 = \frac{t-2}{t}; \quad \ldots; \quad \rho_\tau = \frac{t-\tau}{t}; \ldots \; . \tag{3.16}$$

Namely, the autocorrelation function of an integrated process will typically die out very slowly. In fact, if the initial period for this series was far enough back in time, so that $t >> \tau$, then $\rho_\tau \approx 1$. Essentially, since the only difference between y_t and y_{t-1} is the contemporaneous value of the innovation ε_t, this is not much relatively to all the information previously accumulated from the past innovations.

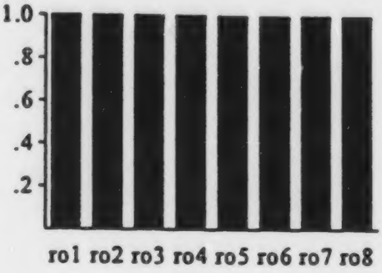


FIGURE 3.1

3.2.4 Practical tools for the identification of stochastic processes

As it is convincingly argued by Granger and Newbold (1986), the identification stage is the most difficult step in the characterization of univariate time series. In fact, as reported half-jokingly by Granger and Newbold (1986), identification requires so much practice and experience to be very good at it that it *"is a technique that should not be attempted for the first time."* More practically speaking, the tools we discuss below can be very helpful in restricting the families of possibly satisfactory models to a few ones. Once the identification stage will have been carried out, a diagnostic stage will follow the estimation and will help us in making the final choice between models that will not have been discarded at first in the identification stage.

Following Box and Jenkins (1970), the basic philosophy of the univariate time series approach is that a simple model reflecting the parsimony principle can often be quite satisfactory to characterize the behaviour of a variable over time. Since any invertible model can be written as an $AR(\infty)$, it could be tempting to include a large number of lagged values of the dependent variable to approximate the underlying data generating process. Such a practice is not advisable since it would likely lead to a model with imprecise parameter estimates. Very often, some simple *ARMA* representation with fairly few parameters can give a good representation.

We now turn to the definitions of practical tools as well as the caveats one must remember when using them for identification.

3.2.4.1 Sample autocorrelations and sample partial autocorrelations

We have discussed what the theoretical ordinary autocorrelation functions and the theoretical partial autocorrelation functions of typical stochastic processes would look like. When working with real data, we then need to estimate the sample counterparts of the theoretical concepts developed above. If we have a sample of T observations, a convergent estimator of the ordinary autocorrelation coefficients is:

$$r_\tau = \frac{\Sigma_{t=\tau+1}^{T}\,(y_t - \bar{y})\,(y_{t-\tau} - \bar{y})}{\Sigma_{t=1}^{T}\,(y_t - \bar{y})^2}, \tag{3.17}$$

where $\bar{y} = T^{-1} \sum_{t=1}^{T} y$. Then, a convergent estimator for the partial autocorrelation coefficient of order τ is given as the solution of the following simultaneous equation system:

$$r_K = \sum_{j=1}^{\tau} \hat{\phi}_{\tau j}\, r_{K-j} \tag{3.18}$$

for K=1, 2, ..., τ. In practice, computer programs often estimate the partial autocorrelation coefficients using Durbin's (1960) computationally efficient recursive algorithm.[11]

As sample counterparts of the ordinary and partial autocorrelations, the estimated values of these coefficients are subject to some degree of imprecision that can be assessed by the standard-errors of r_τ and $\hat{\phi}_{\tau\tau}$. Bartlett (1946) showed that

$$s.e.(r_\tau) \approx T^{-1/2} \cdot \left[1 + 2\,(r_1^2 + r_2^2 + \cdots + r_{\tau-1}^2)\right]^{1/2} \tag{3.19}$$

while Quenouille (1949) showed that

$$s.e.(\hat{\phi}_{\tau\tau}) \approx T^{-1/2}, \tag{3.20}$$

for $\tau > p$ in the case of an $AR(p)$ process. Using these estimated standard errors and relying on asymptotic normality of the innovations, the significance of the sample ordinary and partial autocorrelations is evaluated by a comparison with ±1.96 times the standard errors. Notice that some computer programs do not compute the estimated standard errors of the sample ordinary autocorrelations using equation (3.19), but they instead used equation (3.20). This practice then tends to use estimated values of the standard errors that are too small at low orders for processes containing an autoregressive part. In these instances, we may overstate the significance of the ordinary autocorrelation coefficients.

---

[11] See also Granger and Newbold (1986).

The examination of the sample counterparts to the ordinary autocorrelations and partial autocorrelations can therefore be done with the aim of recognizing behaviours that are compatible with the typical theoretical patterns of specific *ARIMA* processes.

Before turning to other identification tools, it is important to stress one important caveat for applied work. We have seen previously that an integrated series will typically exhibit an ordinary correlogram with a near one first-order autocorrelation, that will die out very slowly even at higher lags. This pattern is therefore expected for the sample autocorrelations of such a series.

In practice, it is worth noticing that with moderate and finite samples of size *T*, the sample correlogram of an integrated series may eventually look as if it is dying for long enough lags, say greater than 10 or 15. This comes from the fact that the sample autocorrelations of an integrated series is expected to die at the rate *T-j/T*. Furthermore, the sample correlogram of some integrated series may die out very slowly but starting with a fairly small value of r_1. For instance, Wichern (1973) has conducted Monte-Carlo experiments for *IMA*(1,1) process, i.e. $y_t - y_{t-1} = \varepsilon_t + \theta_1 \ \varepsilon_{t-1}$. He found that for sample of 50 observations and values of θ_1 being equal to -0.5 and -0.8, the average values of r_1 are 0.62 and 0.21 respectively. Intuitively, this is explained by the fact that even though the underlying stochastic process is nonstationary, as θ_1 approaches of -1, y_t is near stationary.[12] Hence in a finite sample, the correlogram of the process may look like that of a stationary process, as the effect of the unit root will dominate only asymptotically.

We will discuss later formal tests of integration and the use of other tools that may help us confirming or not the need for differencing.

---

[12] For an *IMA*(1,1) process, $y_t = \varepsilon_t$ when $\theta_1 = -1$.

3.2.4.2 Recursive sample autocorrelations

We have so far discussed the role of estimated ordinary and partial correlograms computed for an entire sample in identifying the nature of a stochastic process. This discussion has implicitly assumed that the coefficients of the true underlying ARIMA process, i.e the ϕ's and the θ's, are constant. If this was not the case, the examination of the sample ACF and PACF may be more problematic and may sometimes result in no recognizable and even peculiar and irregular patterns.

A simple technique can be quite helpful in assessing the scope of such problems and was used by Paquet (1989). It consists in computing recursive estimates of the sample ordinary correlogram for different orders and to plot their values over time. The computation of the recursive estimates of the ACF can easily be implemented with the following algorithm:

1. For a full sample of T observations, reckon T sample mean values of the series with $\hat{\mu}(t) \equiv t^{-1} \sum_{j=1}^{t} y_j$ for $t = 1, \ldots, T$.
2. For each $\tau = 1, 2, 3, \ldots,$ compute the T values of $r_\tau(t) \equiv \dfrac{\sum_{j=\tau+1}^{t} (y_j - \hat{\mu}(t))(y_{j-\tau} - \hat{\mu}(t))}{\sum_{j=1}^{t} (y_j - \hat{\mu}(t))^2}$ where $t = 1, \ldots, T$. [13]
3. Plot $r_\tau(t)$ against t for each τ considered.

We should not pay too much attention to the distortions in the recursive estimates caused by too small a sample size in the early part of the sample (say the first 50 or 60 observations). If the stochastic process underlying the series was break free, the recursive sample ACF coefficients should stabilize around their respective true values. If these estimates are not stable, this warns us that it is inappropriate to impose one particular *ARIMA* process with constant parameters to describe a time series over the entire sample.

3.2.4.3 Seasonality and other regular patterns in sample autocorrelations and sample partial autocorrelations

Identification of seasonal and other regular patterns involves further theoretical characterization of ordinary and partial autocorrelation functions that goes beyond the scope of this guide. Such topics are covered in Granger and Newbold (1986). One way to deal with seasonality is, in a first stage, to model the raw unadjusted series as a seasonal *ARIMA*(*Ps*, *Ds*, *Qs*), where s is equal to *4* and *12* for quarterly and monthly data respectively. In a second stage, it remains to identify the *ARIMA*(p, d, q) for the seasonally adjusted series. This would amount to represent a variable y_t with the following stochastic process:

---

[13] This amounts to compute the sample autocorrelation coefficients of the series for increasing subsample by adding up one observation at a time until the whole sample is covered. By comparison, the sample autocorrelations discussed in the previous subsection involved only the use of $r_\tau(T) = r_\tau$.

$$\phi_s(L^s)\,(1-L^s)^D\, y_t = \theta_s(L^s)\, u_t \tag{3.21}$$

where
$$\phi_s(L^s) = 1 - \phi_{1,s}L^s - \phi_{2,s}L^{2s} - \cdots - \phi_{P,s}L^{Ps}$$
$$\theta_s(L^s) = 1 + \theta_{1,s}L^s + \theta_{2,s}L^{2s} + \cdots + \theta_{Q,s}L^{Qs},$$

so that the nonseasonal residual u_t can be thought of as the seasonally adjusted counterpart of y_t. The L^s, i.e. the s^{th} power (and its multiples) of the lag operators in the polynomials above indicates a relationship between obervations s periods apart. Then, the next step consists in characterizing u_t as a regular *ARIMA* such as:

$$\phi(L)\,(1-L)^d\, u_t = \theta(L)\,\varepsilon_t \tag{3.22}$$

where
$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q.$$

This two stage characterization of a time-series amounts to:

$$\phi(L)\,(1-L)^d\,[\phi_s(L^s)\,(1-L^s)^D]\, y_t = \theta(L)\,[\theta_s(L^s)]\,\varepsilon_t\,, \tag{3.23}$$

so that y_t is called an *ARIMA*(p,d,q)(P, D, Q)$_s$.

The issues related to the proper seasonal adjustment can be quite intricate. Finding a multiplicative seasonal ARIMA process to adjust the raw data is one option. There exist also other filters that can serve to adjust the data to remove the contribution of seasonal frequencies (e.g. the U.S. Census X-11 program) that are used by official government agencies that collect and publish time series data. Softwares such as Micro-TSP have a command that carries out some seasonal adjustment in the same spirit as the X-11 procedure.[14]

Sometimes the only data available to work with are published only as seasonally adjusted series. What is important to notice is that as any filters applied to data, automatic seasonal adjustments can create some spurious peaks that sometimes alter the dynamics inherent to a time series. Ideally, it would usually be better to begin empirical works with unadjusted data, which we would learn to characterize ourselves to adjust them properly. In practice, this is very seldomly done. Nevertheless, it is important ro remain aware that the dynamic relationship between macroeconomic variables may be spuriously altered by the presence of both adjusted and unadjusted data in empirical models and that model misspecification as well as poor forecasts can result from adopting officially published seasonally adjusted data.[15, 16]

---

[14] The SEAS command in Micro-TSP computes a centered moving average of the series covering a whole year around the current observation. Then, the multiplicative option computes the ratio of the series to the moving average while the additive option substracts the moving average from the series. The seasonal effects are then reckoned as the average of the ratio associated to each quarterly or monthly observations for all the years. The seasonally adjusted series is finally computed as the original series divided by the multiplicative seasonal effects or as the original series minus the seasonal effects.

3.2.4.4 Hannan and Rissanen's methodology for identification.

As a complemetary tool to the sample ACF and PACF, Hannan and Rissanen (1982) have proposed a procedure that is very helpful to identify the nature of a stochastic process. Let us consider a sample of T observations of a stationary series y_t whose true underlying stochastic process is an $ARMA(p^*,q^*)$.[17]

1. First, approximate the process by some AR(K) process, for different values of K, from 1 to 10. For each value of K, estimate the following equation:
   $y_t = \delta_K + \phi_{K1} y_{t-1} + \phi_{K2} y_{t-2} + \cdots + \phi_{KK} y_{t-K} + \varepsilon_{Kt}$
   and get $\hat{\sigma}_K^2$, i.e. the estimate of the residual variance.
2. For each value of K, compute the Akaike Information Criterion: $\left[\log \hat{\sigma}_K^2 + \frac{2K}{T}\right]$, and choose K^* that minimizes this criterion.[18]
3. Compute the residuals associated with the $AR(K^*)$ approximation: $\hat{\varepsilon}_{K^*t} = y_t - \hat{\phi}_{K^*1} y_{t-1} - \cdots - \hat{\phi}_{K^*K^*} y_{t-K^*}$.
4. Compute the OLS estimator of the model below, for different combinations of p and q:
   $y_t = \delta + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \hat{\varepsilon}_{K^*t-1} + \cdots + \theta_q \hat{\varepsilon}_{K^*t-q} + e_t$
   and get $\hat{\sigma}_{pq}^2 = T^{-1} \Sigma_{t=1}^{T} \hat{e}_t^2$. In practice, you choose $p \leq K^*$, and you can restrict the alternative values of p and q such that $p+q \leq K^*$.

---

[15] Another example of problems associated with the use of officially published seasonally adjusted data is that the adjustment is often carried out on the nominal variables before conversion to real terms. Ghysels and Karangwa (1988) explained that it would be preferable to adjust seasonally data that are first expressed in real terms.

[16] Interested readers on seasonality are referred to Plosser (1979), Ghysels (1990), and Davidson and MacKinnon (1993).

[17] If y_t were $I(d)$, we could differenciate the original series d times to obtain a stationary variable, i.e. $[(1-L)^d y_t]$.

[18] Notice that this stage of the procedure is not consistent and K^* tend asymptotically to overestimate the true order of the model. The second term of the AIC is a penalty for models that contain many regressors.

5. For each considered pair (p, q), compute the following criterion: $\left[\log \hat{\sigma}^2_{p,q} + \frac{(p+q)\log T}{T}\right]$, and choose the 2 or 3 models that minimize this criterion.

This procedure asymptotically has desirable properties and can be used to restrict the class of alternative models to be considered. With moderate sample sizes, Newbold and Bos (1982) investigated the performance of this technique and found it also to be very satisfactory. Once the order of the model has been postulated according to some techniques, we will still have the opportunity to amend the model if the diagnostic tests that will follow the estimation provide evidence of misspecification.

3.2.5 Estimation of a stationary ARMA process

When considering a stationary *ARMA*(p,q) process such as

$$\phi(L)\, y_t = \theta(L)\varepsilon_t \tag{3.24}$$

this can be written as a non-explosive $AR(\infty)$ provided the process is invertible:

$$\varepsilon_t = \frac{\phi(L)}{\theta(L)} y_t = \pi(L) y_t = (1 + \pi_1 L + \pi_2 L^2 + \cdots) y_t\,. \tag{3.25}$$

This can be written also as:

$$\varepsilon_t = (1 + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots + \pi_{t-1} y_1) + \kappa_t \tag{3.26}$$

where the initial condition $\kappa_t = \pi_t y_0 + \pi_{t+1} y_{-1} + \cdots$. It can be shown that if y_t is stationary, this implies that $E(\kappa_t) \to 0$ and $\mathrm{Var}(\kappa_t) \to 0$ as $t \to \infty$.

Since the true innovations are not observable, the *AR* representation defines the innovation in terms of the observable y_t. The π_j parameters are nonlinear functions of the ϕ's and the θ's whose values can be estimated in a way to optimize some criteria. A non-linear least-squares estimator is defined as the solution of

$$\min_{\theta,\phi} \sum_t [(1 + \pi_1 y_{t-1} + \cdots + \pi_{t-1} y_1)^2]. \tag{3.27}$$

An other option is to choose the estimates of the ϕ's and the θ's that maximize the likelihood function of independently normally distributed ε_t's. Actually, the non-linear LS estimator is identical to the MLE estimator conditional upon given initial values for κ_t. The MLE procedure can more generally be computed in estimating as well the initial condition κ_t. This will matter the most if the true process is nearly non-invertible.

Notice finally that the same idea carries for the estimation of more general seasonal *ARMA* models. For further details regarding the estimation of *ARMA* processes, interested readers are referred to Granger and Newbold (1986) or Harvey (1981).

3.2.6 *Diagnotic checking on the residuals of an estimated ARMA process*

Having estimated the retained models, the next stage consists in testing the statistical adequacy of the specification. Essentially, a univariate representation of a time series requires that the residuals of the equations be serially uncorrelated. If this hypothesis is rejected, we must resume our search for a more appropriate model.

3.2.6.1 ACF and PACF of the residuals

One way to do diagnostic checking is to examine the sample ordinary and partial autocorrelation functions of the residuals. Under the null hypothesis that the true error terms of an *ARIMA* equation were white noise, then $r_\tau \sim N(0,\ 1/T)$. This suggests checking the significance at the 5% level of the ACF coefficients by verifying whether they are contained in the interval of $\pm 1.96\ (T)^{-1/2}$. However since we use the estimated residuals and not the true error terms, the standard error of the sample ACF are asymptotically smaller than $(T)^{-1/2}$ for small τ. This is why the conventional procedure would be biased in favour of the nonsignificance of the small order autocorrelations, which could lead us to miss the serial correlation inherent to an inadequate model.

3.2.6.2 Portmanteau statistics

Too many papers use either variants of portmanteau statistics, to test for serial dependence in the residuals of an estimated stochastic process. These statistics are used to test jointly whether the first *M* sample ordinary autocorrelations of the residuals are significant. The most popular versions of these statistics are the Box-Pierce *Q*-statistic,

$$Q_{BP} \equiv T \sum_{\tau=1}^{M} r_\tau^2 \quad \overset{a}{\sim} \quad \chi^2(M), \tag{3.28}$$

and the Ljung-Box *Q*-statistic,

$$Q_{LB} \equiv T(T+2) \sum_{\tau=1}^{M} (T-\tau)^{-1}\, r_\tau^2 \;\overset{a}{\sim}\; \chi^2(M)\,. \tag{3.29}$$

The intuition behind these statistics is that provided that $r_\tau \neq 0$ for some $\tau \leq M$, the computed value of the *Q*-statistics should be big enough so that we would reject the null hypothesis of independently distributed error terms. Unfortunately, these tests generally suffer from an important lack of power, especially in finite samples. This is why their wide spread use is not deserved and other tests, such as the LM test, should be preferred.

3.2.6.3 LM tests of extra coefficients

A preferable and alternative tool for diagnostic checking of the estimated *ARMA*(*p*, *q*) specification is to augment the specification for some values of *p* or *q* greater than that used for the selected model and then to apply a LM test of significance of the added regressors. The test can be constructed as follow:

1. Let the following *ARMA*(*p*, *q*) be the model that was initially estimated over a sample of *T* observations:[19]

   $$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

   Get the estimates of the parameters and the residuals from its estimation, i.e. $\hat{\varepsilon}_t$.
2. Construct the following series:
   $w_t = y_t - \hat{\theta}_1 w_{t-1} - \cdots - \hat{\theta}_q w_{t-q}$,
   $z_t = \hat{\varepsilon}_t - \hat{\theta}_1 z_{t-1} - \cdots - \hat{\theta}_q z_{t-q}$,
   where the lagged values of w_t and z_t in the pre-sample period are set to zero.
3. To test H_0: $y_t \sim ARMA(p, q)$ <=> $\theta_{q+1} = \theta_{q+2} = \ldots = \theta_{q+m} = 0$ against H_1: $y_t \sim ARMA(p, q+m)$, estimate the following equation by OLS:

   $$\begin{aligned}\hat{\varepsilon}_t = {} & \delta + \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \theta_1 z_{t-1} + \cdots \\ & + \theta_q z_{t-m} + \theta_{q+1} z_{t-m-1} + \cdots + \theta_{q+m} z_{t-q-m} + e_t,\end{aligned}$$

   and construct the following *LM* statistics: $T \cdot R^2 \overset{a}{\sim} \chi^2(m)$.

---

[19] The stationary variable y_t can be thought of as being obtained from having differenced enough times an integrated variable.

Although a LM test of the null hypothesis could also be constructed against the alternative hypothesis of an *ARMA*($p+m$, q) with the auxiliary regression being specified as:

$$\hat{\varepsilon}_t = \delta + \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \phi_{p+1} w_{t-p-1} + \cdots$$
$$+ \phi_{p+m} w_{t-p-m} + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q} , \qquad (3.30)$$

the two tests are essentially identical. Therefore, these tests cannot be used to discriminate between adding up extra *AR* parameters or extra *MA* parameters. Such a LM test will help uncover a problem with the maintained specification, but it is in conjunction with an examination of the sample ACF and PACF that we can get some intuition for the amendments that should be brought to the initial model.

It should also be noticed that with *ARMA* specification, residuals-based LM tests of serial correlations that were discussed in the previous chapter are also essentially similar to the LM test for the significance of extra coefficients discussed in this section. Neither one provides extra information relative to the other.

3.2.6.4 Other specification tests

Tests of normality of the innovations or of heteroskedasticity can also be applied when working with *ARMA* processes. They can be constructed in the same manner that was described in chapter 2. For instance, if it was discovered that the variance of the innovations may be varying over time, while the parameters of the process were stable, one should compute White's asymptotically heteroskedasticity-consistent estimator of the variance-covariance matrix of the parameters.

3.2.7 Example: the Canadian exchange rate

3.2.7.1 Identification of the stochastic process. ACF, PACF and other visual tools

Let us consider the logarithm of the Canadian nominal exchange rate, defined as the number of Canadian dollars per US dollar, over the 1973:01-1993:01 period. The raw data for each month, obtained from the Citibase databank, are seasonally unadjusted and are computed as averages of daily figures.

The first panel of Table 3.2 shows the sample ordinary and partial autocorrelation coefficients of the exchange rate level, up to the twelfth order. We refer to these graphs as ACF and PACF respectively. A look at the sample autocorrelation function suggests readily that the variable in level is not stationary and integrated as the coefficients of increasing order die out very slowly. In this case, this suggests differencing the original series before pursuing further the identification stage.[20] Having differenced once the original series, we compute the sample autocorrelations and sample partial autocorrelations that we show in the second panel of Table 3.2. The examination of the ordinary correlogram clearly indicates that there is no need for further differencing. Furthermore, the estimates r1, r8 and r10 are statistically significant at the 5% level. At the same time, the partial autocorrelations at order 1, 8, 10 and 12 are significant at the 5% level. The sample ordinary and partial coefficients at other orders seem not to be different from zero. For the moment, let us ignore the regular pattern in these functions at order 4, 8 and 12, as well as the peak at order 10. The ordinary and the partial correlograms appear therefore to be consistent with either an AR(1), an MA(1) or an ARMA(1,1) stochastic process as the coefficients in either pictures die out very quickly.

To draw inference from the smooth behaviour of the sample autocorrelation function shown in Table 3.2, it is necessary that the stochastic process that generated the series be stable, i.e. it was not affected by any structural break. Such a shift in the data generating process could well be inducing an observed slow decay of the sample autocorrelation coefficients.[21]

---

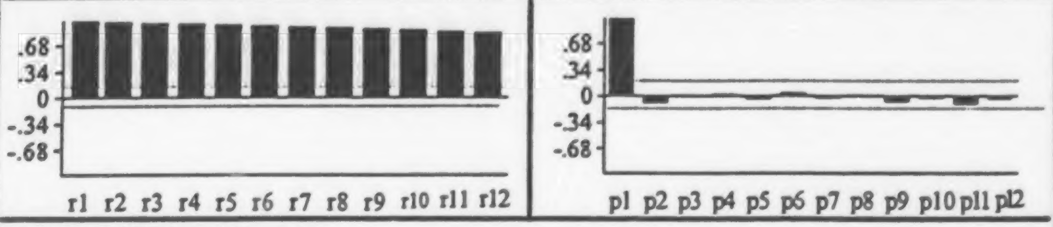
[20] Since the process appears to be nonstationary, the partial autocorrelation function of the variable in level is not really helpful. Essentially, other characteristics of the series are dominated by the evidence of nonstationarity.

[21] As was pointed out by Perron (1988, 1989), changes in mean or breaks in a deterministic trend of a series could lead to spurious evidence favouring the hypothesis of integration when formal tests are performed. Paquet (1989) proposed the examination of plots of the recursive sample ACF to check for instability in the underlying stochastic process.

Table 3.2. - Sample ordinary and partial autocorrelation functions for the Canada-US exchange rate (in log) from 1973:01-1993:01. (241 observations).

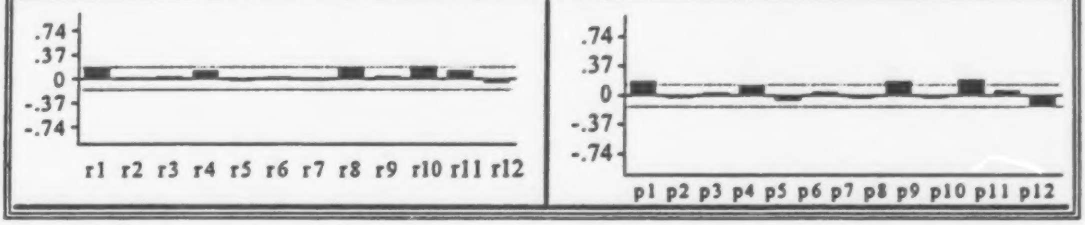
For the level of the series:
($T^{-1/2} = 0,064$)

Sample ACF — Sample PACF



For the first-difference of the series:
($T^{-1/2} = 0,065$)

Sample ACF — Sample PACF



The first panel of Table 3.3 shows plots of the recursive sample autocorrelations at order 1, 2, 3, 4, 5 and 6 for the log-level of the Canadian exchange rate. These graphs do not suggest much instability in the process.[22]

The second panel of Table 3.3 shows plots of the recursive sample autocorrelations, for the first six orders of the first-differenced series, along with a roughly 95% asymptotic confidence band computed as $\pm 1.96/T^{1/2}$.[23] These figures provide essentially the same pattern over the entire sample that was inferred from the examination of the sample ACF in the previous section. We therefore feel confident that we do not make too much offense to the data

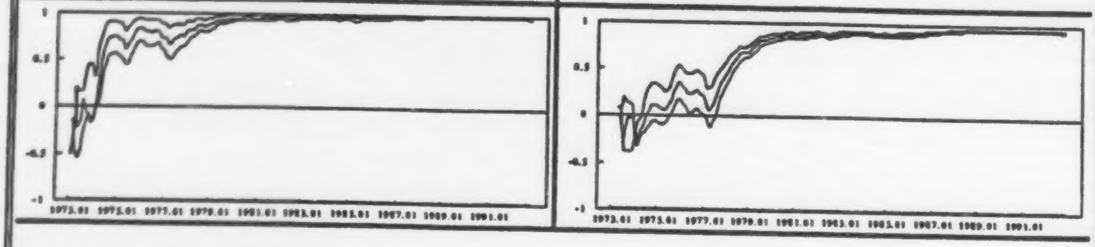
---

[22] Because of the exchange rate crisis following the election of the Parti Québécois in Québec in November 1976, this might suggest some instability of the underlying stochastic process around this period. This might be reflected in the behaviour of the recursive sample ACF between 1976:12 and 1978:06. But since this occurred in the relatively early part of the sample that behaviour could also be due to the relatively small sample size before 1978:06. The possible instability will not be pursued in our example. We will take comfort in the fact that the sample ACF do converge to stable values for most subsamples.

[23] More precisely, the confidence band should have been computed for each t as $\pm 1.96/t^{1/2}$, so that it would be larger in the early part of the sample.

Table 3.3- Recursive sample ordinary autocorrelation function up to order 6 for the Canada-US exchange rate (in log) from 1973:01-1993:01. (241 observations).

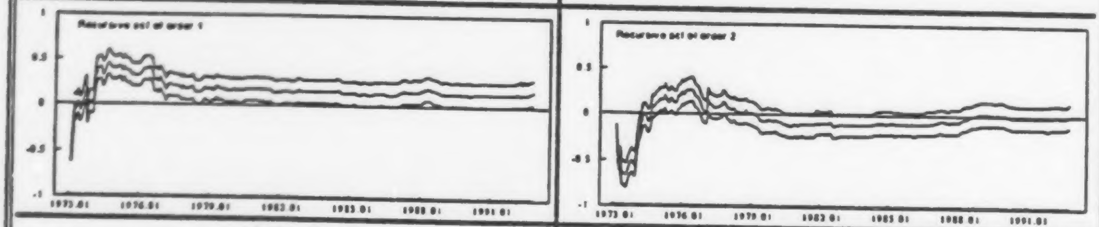
For the level of the series:



Order 1, 2 and 3

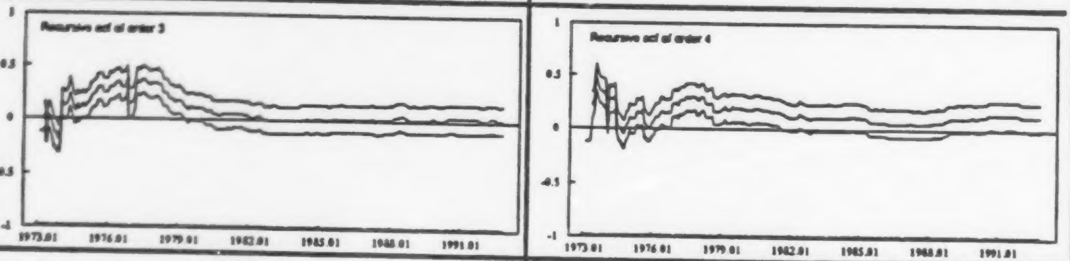
Order 4, 5 and 6

For the first-difference of the series:
(with the asymptotic 95% confidence interval)



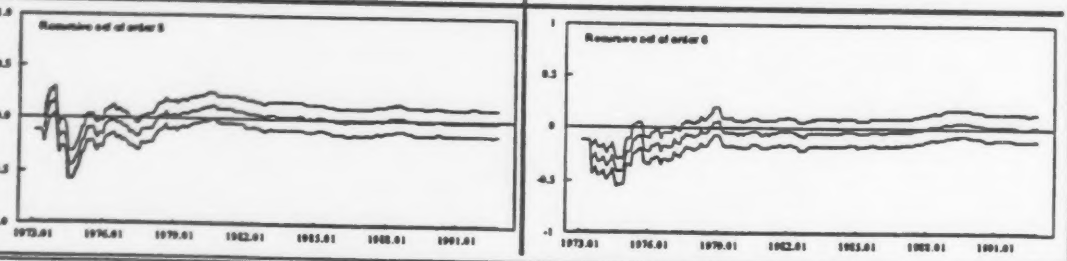
Order 1

Order 2



Order 3

Order 4



Order 5

Order 6

if there is a stable stochastic process underlying the first-differenced logarithm of the Canadian exchange rate.

Referring to the monthly data on exchange rates, the regular pattern of the sample ACF and PACF at lags 4, 8, and 12 would tend to suggest evidence of some seasonal effects. The peak at order 10 and other peaks of the partial autocorrelation function at orders 28 (not shown in the graph above) may be the result of the construction of the data series as monthly averages of daily figures.[24]

Without getting into a formal identification of seasonal and regular pattern for the exchange rate series, there is a satisfactory formulation of a multiplicative *ARMA* polynomial process for our first-differenced series. To capture the seasonal pattern as well as the other particular relationship in the data between observations 4, 8, 10 and 28 months apart, and to take into account the nonstationarity of the level of the variable, we propose a stochastic process of the following form:

$$\phi(L)\,(1-L)\,y_t = \theta(L)\,[(1 + \theta_{4,12}\,L^4 + \theta_{8,12}\,L^8 + \theta_{10,12}\,L^{10} + \theta_{28,12}\,L^{28})]\,\varepsilon_t\ ,$$

where $\phi(L)$ and $\theta(\mathrm{L})$ may be such that the adjusted series u_t is likely to be an $AR(1)$, an $MA(1)$, or an $ARMA(1,1)$.

Even though this is not written exactly with the specification of equation (2.23), it is consistent with the idea embedded in the treatment of repetitive patterns in the observed data some periods apart. We will test later whether it is an appropriate empirical specification.

3.2.7.2 *Identification of its stochastic process. Information criteria analysis.*

Table 3.4 shows the values of the first-stage criterion for different $AR(K)$ approximation of the log- first-differenced Canadian exchange rate. The criterion reaches a local minimum for $K=1$, but seems to be even smaller at $K=10$. Because of the patterns seen in the ACF and the PACF of the series, we would suspect that the lower value is symptomatic of some regularity of the seasonal or of other type. If we decide for the moment to choose $K^* = 1$, then the second criterion favours an $MA(1)$ process.

---

[24] Attempts to adjust seasonally the Canadian nominal exchange rate series with either options of the SEAS command in Micro-TSP do not seem to do a proper job in removing the seasonal pattern of the ACF and the PACF. In fact, the filter described in an earlier footnote tends to add further serial dependence in the series.

Notice that if we constructed a first-differenced log(e_t) series adjusted for the seasonal and atypical serial dependence as

$$\Delta \log(e_t) = (1 + \hat{\theta}_{4,12} L^4 + \hat{\theta}_{8,12} L^8 + \hat{\theta}_{10,12} L^{10} + \hat{\theta}_{28,12} L^{28}) \hat{u}_t ,$$

where u_t is the adjusted series and the coefficients of the polynomials are estimated in a manner discussed below, we can apply Hannan's and Rissanen's (1982) procedure on u_t. This would then suggest unequivocally that the adjusted series may be described as an *MA(1)* process.

Table 3.4. - Akaike Information Criterion for the *AR*(K) approximation of the Canada-US exchange rate (in log-difference) and its adjusted value from 1973:01-1993:01.[1]

K	1	2	3	4	5	6	7	8	9	10
Δlog(e_t)	**-9.19449**	-9.18606	-9.17842	-9.18332	-9.17818	-9.17073	-9.16266	-9.18625	-9.17232	-9.21087
$\hat{u}$	**-9.33465**	-9.3269	-9.32038	-9.31168	-9.3037	-9.29506	-9.29044	-9.28187	-9.27331	-9.26462

[1] A bold number means that the criterion has either reached a local or a global minimum.

Table 3.5. - Information Criterion for different *ARMA*(p,q) of the Canada-US exchange rate (in log-difference) and its adjusted value from 1973:01-1993:01.[1]

p. q	1, 0	0, 1	1, 1
Δlog(e_t)	-9.21345	**-9.21711**	-9.19100
adjusted Δlog(e_t)	-9.24668	**-9.25318**	-9.22528

[1] A bold number means that the criterion has either reached a local or a global minimum.

3.2.7.3 Estimation of a stochastic process.

From the identification described above, the following models have been estimated by non-linear least squares for the 1973:01-1993:01 period:

Model 1:

$$\Delta \log(e_t) = \underset{(0.00064)}{0.001028} + \underset{(0.0638)}{0.1754}\, \varepsilon_{1t-1} + \varepsilon_{1t} \qquad \bar{R}^2 = 0.0265;$$

Model 2:

$$\Delta \log(e_t) = \underset{(0.00061)}{0.00105} + (1 + \underset{(0.0603)}{0.1927} L^4 + \underset{(0.0608)}{0.1702} L^8 + \underset{(0.0602)}{0.1947} L^{10} - \underset{(0.0629)}{0.2783} L^{28}) u_t$$

$$\bar{R}^2 = 0.0993$$

$$u_t = \underset{(0.0633)}{0.1946}\ \varepsilon_{2t-1} + \varepsilon_{2t} \qquad \bar{R}^2 = 0.0355$$

Model 3:

$$\Delta \log(e_t) = \underset{(0.00060)}{0.00105}$$

$$+ \ (1 + \underset{(0.0606)}{0.1992}\, L^4 + \underset{(0.0613)}{0.1819}\, L^8 + \underset{(0.0606)}{0.1737}\, L^{10} - \underset{(0.0635)}{0.2816}\, L^{28})(1 + \underset{(0.0643)}{0.1973}\, L)\, \varepsilon_{3t}$$

$$\bar{R}^2 = 0.1283$$

Model 1 postulates that the first-difference of the log of the nominal Canadian exchange rate follows an *MA*(1) process while ignoring the seasonal and the atypical serial dependence. Model 2 first estimates a specification addressing the seasonality and the relationship in the data between observations 10 and 28 months apart. The residual of this specification is then used as an adjusted series for which we estimate an *MA*(1) model. Finally, model 3 estimates jointly the multiplicative adjusting process and the *MA*(1) behaviour. In the three models, all the slope coefficients are significant at the 1% level. It is interesting to notice that the estimate of the *MA*(1) coefficient is not much affected by the modelling or not of the seasonal and of the atypical serial dependence in the data. Moreover, both the joint and the separate estimation of the multiplicative seasonal and atypical serially dependent processes give similar estimates.

3.2.7.4 Diagnostic Checking: ACF and PACF of the residuals.

Table 3.6. - Sample ACF and PACF for the residuals of different *ARMA* models of the first difference of the Canada-US exchange rate (in log) from 1973:01-1993:01. (240 observations; $1.96/(T^{0.5}) = 0.127$)[1]

Residual	τ	1	2	3	4	5	6	7	8	9	10
ε_{1t}	ACF	0.002	0.010	0.006	**0.130**	-0.045	0.032	-0.047	**0.185**	-0.026	**0.174**
	PACF	0.002	0.010	0.006	**0.130**	-0.046	0.030	-0.049	**0.173**	-0.020	**0.174**
u_t	ACF	**0.183**	-0.004	0.018	-0.028	-0.040	-0.018	-0.067	0.012	0.013	0.012
	PACF	**0.183**	-0.038	0.027	-0.038	-0.028	-0.008	-0.065	0.039	-0.001	0.013
ε_{2t}	ACF	-0.003	-0.008	0.024	-0.026	-0.036	0.002	-0.072	0.024	0.009	-0.002
	PACF	-0.003	-0.008	0.024	-0.026	-0.035	0.001	-0.071	0.025	0.006	0.000
ε_{3t}	ACF	-0.003	-0.007	0.022	-0.031	-0.036	0.002	-0.075	0.015	0.010	0.015
	PACF	-0.003	-0.007	0.022	-0.031	-0.036	0.001	-0.074	0.015	0.006	0.018

[1] A bold number means that the statistic is significant at the 5% level.

An examination of the results shown in Table 3.6 suggests that the residuals of model 1 exhibits seasonal as well as atypical serial dependence. The adjusted variable u_t has ACF and PACF that can be those of a first-order serially correlated stochastic process. While the residuals from the second stage of model 2 and those from model 3 have sample ACF and PACF expected from a white noise process.

3.2.7.5 Diagnostic checking and test statistics.

The Box-Pierce Q-statistics, the Ljung-Box Q-statistics, and their associated P-values are reported in Table 3.7. The P-values indicate the probability under the null hypothesis of having a test statistic whose absolute value is greater than or equal to that of the sample statistic. For instance, with the residuals ε_{1t}, $Q_{BP}(10)$, a P-value of 0.022 means that the null hypothesis that the first 10 sample autocorrelations be jointly equal to zero is rejected at the 3% but not at the 2% level. The lower power of these test statistics is suggested when looking at the Q-statistics of u_t at order 4. The test would not reject the null hypothesis at a 5% significance level, even though Table 3.6 indicated a significant first-order sample autocorrelation for u_t.

Tables 3.8 and 3.9 report further diagnostic tests on models 1 and 3.[25] While additional MA coefficients are significant and lead to the rejection of the specification maintained in model 1, the LM test for extra lags does not provide evidence at a 96% significance level, against model 3. The tests reported in Table 3.9 provide no significant evidence against either the normality of the error terms, or the specifications of the conditional mean represented by the chosen *ARIMA* of model 1 and 3, at conventional statistical levels. However, the ARCH test at order 4 suggests that the innovations of model 1 may be heteroskedastic. Given that the sample ACF and PACF reported in Table 3.6 and the LM test reported in Table 3.8 supported a richer model, this evidence could be symptomatic of the same problem. In fact, the ARCH test conducted on the residuals of model 3 do not suggest any autoregressive conditional heteroskedasticity.

Hence, from the identification, the estimation, and the diagnostic checks, model 3 appears to be a satisfactory representation of the Canadian nominal exchange rate series under consideration.

---

[25] Models 2 and 3 being equivalent we chose not to duplicate similar diagnostic tests on ε_{2t} and ε_{3t} .

Table 3.7. - Box-Pierce and Ljung-Box Q-statistics and their p-values for the residuals of different *ARMA* models of the first difference of the exchange rate (in log) from 1973:01-1993:01. (240 observations)[1]

Residual	M	4		8		10		12	
		Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
ε_{1t}	Q_{BP}	4.10	0.393	13.55	0.094	**20.95**	**0.022**	**23.76**	**0.022**
	Q_{LB}	4.20	0.380	14.05	0.081	**21.83**	**0.016**	**24.81**	**0.016**
u_t	Q_{BP}	8.33	0.080	9.91	0.271	9.98	0.442	11.29	0.505
	Q_{LB}	8.44	0.077	10.08	0.260	10.15	0.427	11.53	0.484
ε_{2t}	Q_{BP}	0.32	0.989	2.00	0.981	2.02	0.996	3.41	0.992
	Q_{LB}	0.32	0.988	2.07	0.979	2.09	0.996	3.56	0.990
ε_{3t}	Q_{BP}	0.37	0.985	2.00	0.979	2.14	0.995	3.41	0.992
	Q_{LB}	0.37	0.985	2.13	0.977	2.22	0.994	3.56	0.990

[1] A bold number means that the statistic is significant at the 5% level.

Table 3.8. - LM test of H_0: *ARMA*(p, q) versus H_1: *ARMA*$(p, q+m)$ and their p-values for the first difference of the exchange rate (in log) from 1973:01-1993:01. (240 observations)[1]

Model	m	4		8		10		12	
		Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
1	*LM*	4.55	0.208	12.78	0.078	17.69	0.039	23.06	0.017
3	*LM*	0.23	0.973	1.97	0.962	1.99	0.992	2.92	0.992

[1] A bold number means that the statistic is significant at the 5% level.

Table 3.9. - Other diagnostic tests and their p-values for models of the first difference of the exchange rate (in log) from 1973:01-1993:01. (240 observations)[1]

Model	RESET(3)		ARCH(1)		ARCH(4)		NORMALITY	
	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
1	0.83	0.842	4.20	0.040	7.04	0.134	1.58	0.454
3	0.671	0.880	0.512	0.474	6.28	0.179	1.75	0.417

[1] A bold number means that the statistic is significant at the 5% level.

3.3 The Issue of Nonstationarity and Unit Roots in Macroeconomics Time Series.

3.3.1 Introduction

One particular type of nonstationarity in time series is the integration of a variable. That is, an integrated series of order *d*, denoted $I(d)$, is such that it can be represented by a stationary and invertible *ARMA* process once it has been differenced *d* times. Hence, the variable $y_t \sim I(d)$, if $[(1-L)^d y_t] \sim ARMA(p, q)$ that is stationary and invertible. It is also said that the variable y_t has *d* unit roots since the *d* roots of the polynomial $(1-L)^d$ are equal to *1*. The search for unit roots in time series, and the statistical methods to deal with integrated variables have constituted one of the most prolific areas of research in econometrics, and in macroeconomics in the 1980's.

We have described previously what would look like the ACF of an integrated series. In this section, we will discuss the categorisation of trends in time series as either deterministic or stochastic, the definitions and uses of unit root tests, and the relative relevance of unit roots in macroeconomics and econometrics.

3.3.2 Deterministic versus stochastic trend

In a seminal paper, Nelson and Plosser (1982) have proposed to distinguish time series from the fundamental nature of their underlying trend. Some time series should be best described as having a deterministic trend while others should rather be described as having a stochastic trend. The fundamental difference between the two is that the former is such that the time series is mean-reverting while the latter will have no such inherent tendency.

More precisely, a time series, y_t, that exhibits a deterministic trend can be represented as:

$$y_t = \delta + \gamma\, t + u_t, \tag{3.31}$$

where $\phi(L)\, u_t = \theta(L)\, \varepsilon_t$, $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$, and $\phi(L)$ and $\theta(L)$ are respectively stationary and invertible. Such a series has an unconditional (or long-run) mean equal to $E(y_t) = \delta + \gamma\, t$, i.e. the deterministic trend of the series. The long-run forecast error of y_t is given by u_t that is stationary.[25] Hence, there is no influence of past or current innovations on the long-run mean of the variable, and all uncertainty about the long-term forecast error of y_t is bounded. A series that exhibits a deterministic trend is also called a trend-stationary time series. Panel A in Table 3.10 shows the evolution of a variable y_t that has a deterministic linear time trend.

---

[25] Notice that taking the first difference of y_t will yield a process with an non-invertible component in its *MA* part: $y_t - y_{t-1} \equiv \Delta y_t = \gamma + (1-L)\, u_t$.

By contrast, a time series that is characterized as having a stochastic trend will be of the type:

$$y_t = \gamma + y_{t-1} + u_t \,, \tag{3.32}$$

where $\phi^*(L)\,u_t = \theta^*(L)\,\varepsilon_t$, $\varepsilon_t \sim i.i.d.(0, \sigma^2_\varepsilon)$, and $\phi^*(L)$ and $\theta^*(L)$ are respectively stationary and invertible. This can be written also as:

$$(1-L)\,y_t = \gamma + u_t \,, \tag{3.33}$$

which amounts to:

$$y_t = y_0 + \gamma\, t + \sum_{j=1}^{t} u_{t-j} + u_t \,. \tag{3.34}$$

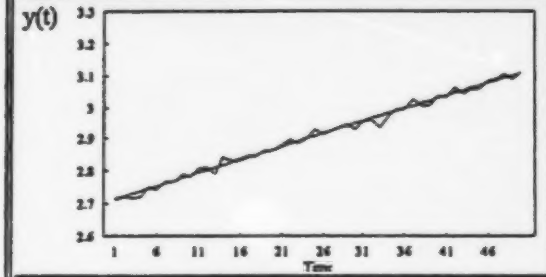
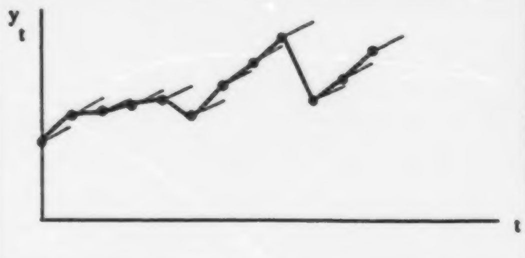
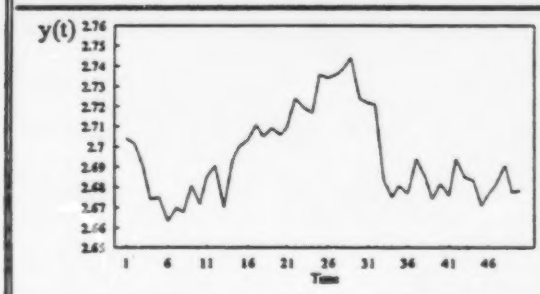
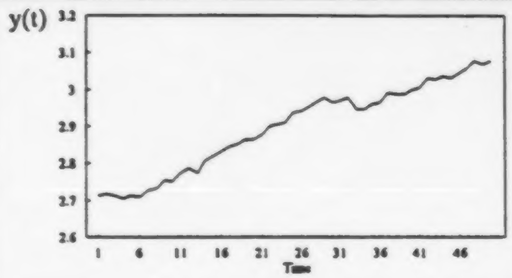
What is typical of such a process is that the intercept of y_t is function of historical events as given by the initial value of the process y_0, and the value taken for the variable at any point t is function of all past and current innovations. Also, the long-run variance of y_t is given by:

$$Var(y_t) = t \cdot Var(u_t) \rightarrow \infty \quad \text{as} \quad t \rightarrow \infty \,, \tag{3.35}$$

i.e. the $Var(y_t)$ explodes as time goes by, and y_t has no tendency to return to the deterministic trend. In fact, if we were attempting to make the series stationary by removing the influence of the deterministic trend component [$\gamma\, t$], we would still be left with a non-stationary time series. That is since deviations about the deterministic trend are the accumulation of all past and current stationary innovations [$\Sigma^t_{j=0}\, u_{t-j}$].

Panel B in Table 3.10 shows a schematic representation of a time series that has a stochastic trend with u_t being equal to ε_t that is a white noise. From each point in time $t\text{-}1$, the best forecast one can make of the series for next period is given by [γ + y_{t-1}]. But, the realisation of the current innovation u_t makes you deviate from this forecast at the same time that it will shift the starting point from which you will formulate your best forecast for period $t+1$. This illustrates why the detrending method employed in applied research is not innocuous.

Table 3.10. - Representations of time series generated by different stochastic processes [with $\varepsilon_t \sim$i.i.d. $N(0, 0.011^2)$]

A) A variable with stationary fluctuations around a deterministic trend:	B) Schematic representation of a stochastic trend:
$y_t = 2.70805 + 0.008 \cdot t + \varepsilon_t,$	
	
C) A non-stationary variable: a random walk:	D) A variable with a stochastic trend: a random walk with drift:
$y_t = y_{t-1} + \varepsilon_t,$ with $y_0 = 2.70805$	$y_t = 0.008 + y_{t-1} + \varepsilon_t,$ with $y_0 = 2.70805$
	

We can see from equation (3.33) that a series y_t with a stochastic trend has (at least) one unit root in the autoregressive component of the data generating process of its level. To detrend such a time series, it is therefore necessary to take the first-difference of the variable.[26] Such a variable y_t is said to be difference-stationary.

When u_t is white-noise, the stochastic trend process represented by equation (3.32) can specialize to the case of a pure random walk if $\gamma = 0$, and of a pure random walk with drift if $\gamma \neq 0$. Panels C and D in Table 3.10 illustrate what such processes could look like in a particular example. Notice that because of the initial value and the value of γ that was chosen in the illustrative examples, we can see how the presence (panel D) or not (panel A) of the unit root can influence the evolution of the series over time.

---

[26] If the series were $I(d)$, it would be necessary to take the d^{th}-difference of the variable. In practice, it is rather seldom that we encounter data with d being different from 1 or, sometimes, 2.

3.3.3 Unit root tests.

3.3.3.1 Cochrane's variance ratio test

We have already discussed that a lengthy decline of the sample autocorrelation function can suggest an integrated process as it reflects the long influence an innovation that occurred even very far in the past can still exert in the current level of a time-series. Cochrane (1988) has proposed another tool to assess this source of nonstationarity.

If one considers some stationary $I(0)$ process y_t, then $var(y_{t+1}-y_t) = 2\ var(y_t) + c_1$, and $var(y_{t+k+1}-y_t) = 2\ var(y_t) + c_k$, where c_1 and c_k are nonnegative constants.[27] While if $y_t \sim I(1)$, $var(y_{t+k+1}-y_t) = (k+1)\ var(y_{t+1}-y_t)$. Consequently, the following ratio of variances should provide indication of nonstationarity:

$$V^k \equiv \frac{1}{k+1}\frac{var\,(y_{t+k+1}-y_t)}{var(y_{t+1}-y_t)} = 1 + 2\sum_{j=1}^{k}\left(1-\frac{j}{k+1}\right){}_{\Delta y}\rho_j \quad , \qquad (3.36)$$

where ${}_{\Delta y}\rho_j$ is the j^{th}-order autocorrelation coefficient for the first-difference of y_t.

For a driftless integrated process with white noise innovation, the ratio V^k tends to 1 as k approaches ∞. If the innovation is serially correlated or if the integrated process has a drift, then V^k tends to some constant greater than 0, as $k \rightarrow \infty$. However, if the process is stationary, then V^k tends to 0 as k increases. This can be understood as the first-difference of a trend-stationary variable should have negative autocorrelation coefficients at some lags that cancel out the positive coefficients.

The sample analogue to V^k and its estimated standard error can be computed as:[28]

$$\hat{V}^k = 1 + 2\sum_{j=1}^{k}\left(1-\frac{j}{k+1}\right){}_{\Delta y}r_j \quad \text{and} \quad S.\hat{E}.(\hat{V}^k) = \frac{\hat{V}^k}{\sqrt{\frac{3}{4}(T/(k+1))}} . \qquad (3.37)$$

In practice with moderate sample sizes, one must be careful not to compute Cochrane's ratio with too large values of k. Generally, the evolution of Cochrane's ratio is fairly easy to characterize with increasing k up to half the sample size. The estimate of the standard-error of the variance ratio suggests also that too large a k relative to the sample size yields higher standard-errors of the estimated V^*, as there is less information to reckon the higher orders autocorrelations of the first-differenced series. This is equivalent to the issue of very low precision one can obtain with existing sample sizes to infer the contribution of the very long period (or short-frequency) to the variance of the change in a time series.

---

[27] If y_t is white noise, c_1 and c_k are zero. More generally, if y_t is stationary and ergodic then $cov(y_{t-k}, y_t) \rightarrow 0$ as $k \rightarrow \infty$.

[28] The Cochrane ratio statistics can be obtained as the Bartlett window estimate of the spectral density function of Δy_t evaluated at the zero frequency. The interested reader in spectral analysis is referred to Granger and Newbold (1986), Kennedy (1992) and Paquet (1993). Sheffrin (1989) provides an intuitive and heuristic interpretation of Cochrane's variance ratio statistic, in the frequency domain.

3.3.3.2 Simple and augmented Dickey-Fuller tests

Discussion of the issues relevant to unit root tests can be most effectively conducted by focusing on the three following alternative data generating processes (DGP) of a time-series:[29]

$$y_t = \rho_n y_{t-1} + u_t \tag{3.38}$$

$$y_t = \mu_c + \rho_c y_{t-1} + u_t \tag{3.39}$$

$$y_t = \mu_{c\tau} + \gamma_{c\tau} t + \rho_{c\tau} y_{t-1} + u_t \tag{3.40}$$

where $\phi(L) u_t = \theta(L) \varepsilon_t$, and $\varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2)$. Let us first assume that u_t be equal to ε_t so that it is a white noise, under some specific values of the parameters, these processes can be characterized in the following manners.

- For equation (3.38):
  - if $\rho_n < 1$, then the DGP is a stationary zero-mean first-order autoregressive process;
  - if $\rho_n = 1$, then the DGP is a pure (driftless) random walk, i.e. non-stationary.
- For equation (3.39):
  - if $\rho_c < 1$, then the DGP is a stationary first-order autoregressive process, with mean $\mu_c/1-\rho_c$;
  - if $\rho_c = 1$, then the DGP is a random walk with drift μ_c, i.e. non-stationary.
- For equation (3.40):
  - if $\rho_{c\tau} < 1$, then the DGP is a trend-stationary first-order autoregressive process, with mean $\frac{\mu_{c\tau}}{1-\rho_{c\tau}} + \gamma_{c\tau} \Sigma_{j=0}^{t} \left[\rho_{c\tau}^{j} \cdot (t-j) \right]$;
  - if $\rho_{c\tau} = 1$, then the DGP is a random walk with a drift changing over time.

---

[29] In equation (3.40), the linear time trend variable t could be replaced by a centred time trend defined as $[t- T/2]$ with no consequences for the testing procedures. In fact, this was the definition employed for instance in Perron (1988, 1989), and in Phillips and Perron (1988) papers, which will be discussed later.

Interestingly, equation (3.40) can also be given another interpretation for which it nests the two different trend representations we discussed in the previous section. Indeed, for the trend-stationary model:

$$y_t = \delta + \gamma_c\, t + u_{ct}\,, \tag{3.41}$$

and the difference-stationary model (when $\rho_c = 1$):

$$y_t = \gamma_c + \rho_c\, y_{t-1} + u_{ct}\,, \tag{3.42}$$

equation (3.40) can be interpreted as the reparameterization of the following general non-linear model that specializes to either (3.11) or (3.42) when either $|\rho| < 1$ or $|\rho| = 1$, respectively:

$$\begin{aligned} y_t &= \delta + \gamma_c t + \rho_c [y_{t-1} - \delta - \gamma_c \cdot (t-1)] + u_{ct} \\ &= [\delta(1-\rho_c) + \gamma_c \rho_c] + [\gamma_c(1-\rho_c)]\, t + \rho_c y_{t-1} + u_{ct} \\ &= \mu_{c\tau} + \gamma_{c\tau}\, t + \rho_{c\tau}\, y_{t-1} + u_{c\tau\, t}\,, \end{aligned} \tag{3.43}$$

where $\rho_{c\tau} = \rho_c$ and $u_{c\tau\,t} = u_{c\,t}$. Under this interpretation, it is important to notice that $\gamma_{c\tau} = 0$ whenever $\rho_{c\tau} \equiv \rho_c = 1$. Hence, equation (3.40) collapses to the DGP represented by equation (3.39) under the unit root hypothesis. This plays a crucial role in performing and interpreting correctly unit root tests based on the empirical specifications representing the appropriate DGP (3.38), (3.39) or (3.40) under the null hypothesis of integration.

To account for these different DGPs, the empirical specification to be estimated must nest both the null and the alternative hypotheses. Hence, we present below the different specifications and the null hypothesis (under $\rho = 1$) and the alternative hypotheses (under $\rho < 1$) they admit.

Model (A):

$$y_t = \rho_n\, y_{t-1} + u_t$$

$$\Leftrightarrow \qquad \Delta y_t = (\rho_n - 1)\, y_{t-1} + u_t, \tag{3.44a}$$

to test H_0 of a pure random walk with initial condition y_0 being 0 against H_1 of a zero-mean *AR*(1) process.

Model (B):

$$y_t = \beta_{0c} + \rho_c y_{t-1} + u_t$$

$$\Leftrightarrow \qquad \Delta y_t = \beta_{0c} + (\rho_c - 1)\, y_{t-1} + u_t \ , \tag{3.44b}$$

to test H_0 of a pure random walk with initial condition y_0 being 0 against H_1 of a non-zero-mean *AR*(1) process.

Model (C):

$$y_t = \beta_{0c\tau} + \beta_{1c\tau} t + \rho_{c\tau} y_{t-1} + u_t$$

$$\Leftrightarrow \qquad \Delta y_t = \beta_{0c\tau} + \beta_{1c\tau} t + (\rho_{c\tau} - 1)\, y_{t-1} + u_t \ , \tag{3.44c}$$

to test H_0 of a random walk with drift against H_1 of a linear trend-stationary *AR*(1) process.

Model (D):

$$y_t = \beta_{0c\tau\tau} + \beta_{1c\tau\tau} t + \beta_{2c\tau\tau} t^2 + \rho_{c\tau\tau} y_{t-1} + u_t$$

$$\Leftrightarrow \Delta y_t = \beta_{0c\tau\tau} + \beta_{1c\tau\tau} t + \beta_{2c\tau\tau} t^2 + (\rho_{c\tau\tau} - 1)\, y_{t-1} + u_t \ , \tag{3.44d}$$

to test H_0 of a random walk with time-varying drift against H_1 of a quadratic trend-stationary *AR*(1) process.

In all cases above, both representations of each model are equivalent reparameterization and can be used in empirical work. What should be emphasised however is that under the null hypothesis of $\rho = 1$, β_{0c} in equation (3.44b), $\beta_{1c\tau}$ in regression (3.44c) and $\beta_{2c\tau\tau}$ in regression (3.44d) must equal zero. Under the alternative, these coefficients would be different from zero. Hence, regression (3.44a) should be of use only in the case of a potentially integrated series whose mean is zero.[30] Regression (3.44b) will be useful to test for a unit root in the case of a driftless integrated process with non-zero mean associated to the DGP (3.38). Under the null hypothesis, regression (3.44c) will correspond to that of an integrated process with drift (3.39). Finally, regression (3.44d) should be used to uncover evidence favouring the DGP (3.40) under integration with time-shifting drift.[31] The latter has not been much considered in applied work.

---

[30] As we will discuss in the next chapter, this specification will be of use in residuals based tests of no cointegration, provided that any relevant deterministic regressors (e.g. mean, linear, or quadratic trend) were included in the regression equation with integrated variables. These are unit root tests applied on residuals.

[31] The appropriate use of the right regression model to begin with depends also technically on the necessity that the order of the process y_t, that determines the rates of convergence of its moments, be the same under the null hypothesis of integration and under the alternative of stationarity. See Davidson and MacKinnon (1993) for further discussions on this topic.

In their seminal papers, Dickey (1976), Fuller (1976) and Dickey and Fuller (1981) have originally developed unit root test statistics for regressions (3.44a) to (3.44c), for the case of a DGP with $\rho = 1$, under the maintained assumptions that $y_0 = 0$, and that $u_t = \varepsilon_t \sim$i.i.d.$(0, \sigma_\varepsilon^2)$. In either cases, Dickey and Fuller proposed two alternative statistics of the single hypothesis H_0: $\rho = 1$ versus H_1: $\rho < 1$. These are presented in Table 3.11 along with other tests that will be discussed shortly.[32] First, one can compute a familiar *t*-type statistics defined as:

$$t_\rho = \frac{\hat{\rho} - 1}{\hat{se}\,(\hat{\rho})}. \tag{3.45}$$

The practical advantage of estimating the second empirical specifications is that the computed value of t_ρ is printed out in output regressions besides the estimated value of the coefficient (ρ-1). The alternative statistic is the so called normalized bias test defined as:

$$T(\hat{\rho} - 1). \tag{3.46}$$

In contrast to the usual *t*-statistics, under the null hypothesis of nonstationarity (i.e. $\rho = 1$), t_ρ does not follow the Student-*t* distribution even asymptotically, but rather follows a non-standard Dickey-Fuller distribution (see Table A3.1). Similarly, the normalized bias statistic does not follow a standard-normal distribution (see Table A3.2). This is true even though it has been shown that the OLS estimate of ρ is superconsistent towards its true value under the null hypothesis.[33] In fact, if we were to compare mistakenly the computed values of these statistics with critical values picked from the usual tables, we would be rejecting too often the null hypothesis of a unit root. The actual Dickey-Fuller distribution for t_ρ is skewed to the left compared with the normal distribution. Dickey and Fuller tabulated the critical values of these test statistics from Monte-Carlo simulations, for only a few finite sample sizes on at most 10 000 replications. MacKinnon (1991), and Davidson and MacKinnon (1993) have recently computed differently the critical values of these statistics in a way that may be more accurate (see Tables A3.3 and A3.4).[34, 35]

---

[32] It is possible to conduct the test against an alternative that includes the possibility of an explosive root, i.e. $\rho > 1$. In most applied work, it is more usual to test the null against the stationary alternative. That is why only the critical values useful against the latter have been reported in tables enclosed. The critical values pertaining to the other side of the distribution can be found in the original published papers as well as in Banerjee *et al.* (1993)

[33] Phillips and Durlauf (1986), and Stock (1987) have shown that the OLS estimator of the coefficient of an integrated series converges at rate T rather then at rate $T^{1/2}$. The derivation of the asymptotic distribution of the OLS estimator requires however the use of functional central limit theorems rather than the usual central limit theorems presented in basic statistics courses.

It is also important to notice that the asymptotic and finite distributions of the t_ρ statistic and of the normalized bias statistic are dependent upon the set of exogenous or deterministic regressors (such as an intercept or a time trend) being included in the regression. This is why both Tables A3.3 and A3.4 of critical values contain different panels which correspond to the regression model (3.44a) without a constant, (3.44b) with a constant without trend, and (3.44c) with both a constant and a trend under the null hypothesis.[36, 37] In the case of equation (3.44d), asymptotic values for these tests are reported in Table A3.4.

Distributions of other single hypothesis *t*-tests and of F-type tests for joint hypotheses relevant to the empirical specifications above have been developed in the literature. Regarding specification (3.44b), one can construct a *t*-test of H_0: $\beta_{0c} = 0$, and an F-test, Φ_1, of H_0: $\beta_{0c} = 0$ and $\rho_c = 1$. With respect to specification (3.44c), the $t_{0c\tau}$ and $t_{1c\tau}$ statistics test respectively that H_0: $\beta_{0c\tau} = 0$, and H_0: $\beta_{1c\tau} = 0$, while the Φ_2-statistic tests H_0: $\beta_{0c\tau} = 0$, $\beta_{1c\tau} = 0$ and $\rho_{c\tau} = 1$. The Φ_3-statistic tests H_0: $\beta_{1c\tau} = 0$ and $\rho_{c\tau} = 1$. However, under the null hypothesis of a unit root, the coefficients on the intercept or on a deterministic time trend, and test statistics about them do not follow standard distributions. Dickey and Fuller (1981) have shown that the distribution of the *t*-statistic to test the nullity of either the intercept, or the linear trend is bimodal, and is more spread than the Student's *t* distribution. The formulae to compute these statistics and references to their critical values, which are reported in this chapter appendix, are given in Table 3.11. We will come back later to a testing strategy that can help to discriminate between the different DGPs.

---

[34] MacKinnon (1991) computed the critical values of the tests, from a more extensive set of simulations, which he summarized with response surfaces that relate the exact critical values of the tests to a constant term (the asymptotic critical value) and terms involving sample size.

[35] To compute the finite sample critical values at a *p*% significance level one only needs to use the estimates of φ_0, φ_1, and φ_2, reported in Table A3.3, that are associated with the relevant significance level, and to evaluate the expression: $C(p) = \varphi_0 + \varphi_1 T^{-1} + \varphi_2 T^{-2}$.

[36] Evans and Savin (1981) and Nankervis and Savin (1985) report Monte Carlo experiments that also illustrate how a non-zero value of the initial condition y_0 can bias the test statistics towards the non-rejection of the unit-root hypothesis. This is even if the asymptotic behaviour of unit root tests is not affected by the value of y_0 under some regularity conditions. For these reasons, one should avoid interpreting statistical evidence based only on equation (3.44a) as supporting the unit root hypothesis if no prior analysis has been conducted with the more general specification.

[37] There is another reason to emphasize the caution not to base unit root tests on regression (3.44b) if one suspects that the underlying stochastic process of the series is integrated with a drift. In this case, unit root tests have virtually no asymptotic power against the null hypothesis. This is why even if West (1988a) has shown that the asymptotic limiting distribution of unit root tests based on equation (3.45) is normal when y_t is integrated with drift, this is of no practical help since, in finite samples, the Dickey-Fuller distributions are more appropriate.

One problem with the tests exposited above is that their critical values have been derived for a process that assumes $y_0 = 0$ and that the innovations of the process are white noise. Although some regularity conditions may assure that the unit root tests are not asymptotically affected by the value of y_0, in practice, one is very likely to encounter data whose innovations exhibit all sorts of dynamics best represented by some *ARMA* process. Dickey and Fuller (1979) and Said and Dickey (1984) have shown that to account for the time-dependence in the innovation, one can augment the basic regression models (3.44a) to (3.44d) with, say, l lags of Δy_t. Then, one computes the t_ρ statistic, which follows the same asymptotic distribution as that reported in the case where u_t is a white noise. This is called the Augmented Dickey-Fuller (ADF) test.[38]

$$\Delta y_t = (\rho_n - 1)\, y_{t-1} + \sum_{i=1}^{l} \alpha_i\, \Delta y_{t-i} + v_t \tag{3.47a}$$

$$\Delta y_t = \beta_{0c} + (\rho_c - 1)\, y_{t-1} + \sum_{i=1}^{l} \alpha_i\, \Delta y_{t-i} + v_t \tag{3.47b}$$

$$\Delta y_t = \beta_{0ct} + \beta_{1ct}\, t + (\rho_{ct} - 1)\, y_{t-1} + \sum_{i=1}^{l} \alpha_i\, \Delta y_{t-i} + v_t \tag{3.47c}$$

$$\Delta y_t = \beta_{0ctt} + \beta_{1ctt}\, t + \beta_{2ctt}\, t^2 + (\rho_{ctt} - 1) y_{t-1} + \sum_{i=1}^{l} \alpha_i\, \Delta y_{t-i} + v_t \tag{3.47d}$$

In fact, if u_t followed an *AR* process with a known order p, then l should be set to the value of p. If u_t follows some *ARMA* process, technically it could be approximated by an $AR(\infty)$. In practice, Said and Dickey (1984) showed that it would be sufficient to increase the number of lags on Δy_t as a function of the sample size, at a rate no faster than $T^{1/3}$. This does not suggest, however, what should be the appropriate value of l in a given application. This is no trivial matter. Choosing too small an l will distort the size (significance level) of the statistics. Choosing too big an l will reduce the degrees of freedom and the power of the test, which will too often lead falsely to the non-rejection of the unit root hypothesis.

Campbell and Perron (1991) suggested that one uses a data dependent procedure to select the value of l:

---

[38] Although it is also possible to compute an analogue to the normalized bias test based on the augmented regression in the case where u_t follows a finite *AR* process, this is not so easy to implement and it is almost never used in practice. See Dickey, Bell and Miller (1986).

1. Start with some upper bound on l, say l_{max}, estimate the augmented regression and check the significance of $\Delta y_{t-l_{max}}$ using the standard normal distribution at some level, say at 10%.
2. If it is significant, proceed with the unit root test. If not, reestimate the augmented regression after having reduced l by one and check the significance of the last included lag on Δy_t.
3. Repeat step 2, until you obtain a significant coefficient on the last included lag on Δy_t, or until you reach $l = 0$. Then, proceed with the unit root test.

Since the maximum lag length, l_{max}, is selected *a priori*, the results of unit root tests may still be sensitive to it. Ghysels and Perron (1993) suggest in fact that the value of l_{max} be chosen beyond the seasonal lag. As they show, this is especially important when working with officially seasonally adjusted time-series. It might also be informative to assess the robustness of the results for several possible values of l_{max}, as in Perron (1993).

Another possibility is to select the lag truncation parameter that minimizes some information criterion such as Akaike's (*AIC*) or, as suggested by Hostland (1992), the *Finite Prediction Error* (*FPE*) criterion. Ng and Perron (1993) find that these information criteria tend to select models that are too parsimonious, which may lead to serious size distortions in unit root tests. They recommend instead the use of Campbell and Perron's (1991) recursive *t*-statistic procedure on the last lag of Δy_t, which leads to statistics with more stable statistical sizes and higher power.

The augmented Dickey-Fuller procedure attempts to account parametrically for the serial correlation in the error terms. We now turn to an alternative procedure based on a nonparametric approach to unit root tests.

Table 3.11.- Test statistics of unit roots and of other hypotheses in unit roots regressions.[1]

Null hypothesis	Comment	Test statistics	Critical values
Model (A): $y_t = \rho_n y_{t-1} + u_t$			
$H_0: \rho_n = 1$	Simple Dickey-Fuller for $y_0 = 0$, and u_t being white noise	$t_{\rho n} = \frac{(\hat{\rho}_n - 1)}{\hat{se}\,(\hat{\rho}_n)}$	Tables A3.1, A3.3, or A3.4
	Normalized bias test for $y_0 = 0$, and u_t being white noise	$T(\hat{\rho}_n - 1)$	Tables A3.2 or A3.4
	Augmented Dickey-Fuller tests for $y_0 = 0$, and u_t being some *ARMA(p,q)*	Based on: $\Delta y_t = (\rho_n - 1) y_{t-1} + \sum_{i=1}^{l} \alpha_i \Delta y_{t-i} + v_t$	
		$t_{\rho n} = \frac{(\hat{\rho}_n - 1)}{\hat{se}\,(\hat{\rho}_n)}$	Tables A3.1 A3.3 or A3.4 asymp
	Perron-Phillips t-test for $y_0 = 0$, and u_t being some *ARMA(p,q)*	$Z(t_{\rho n}) = \left(\frac{S_u}{S_{Tl}}\right) t_{\rho n} - \frac{1}{2}\left(S_{Tl}^2 - S_u^2\right)\left[S_{Tl}\left(T^{-2}\sum_{i=2}^{T} y_{t-i}^2\right)^{1/2}\right]^{-1}$	Tables A3.1, A3.3 or A3.4 asymp.
	Perron-Phillips normalized bias test for $y_0 = 0$, and u_t being some *ARMA(p,q)*	$Z(\hat{\rho}_n) = T(\hat{\rho}_n - 1) - \frac{1}{2}\left(S_{Tl}^2 - S_u^2\right)\left[T^{-2}\sum_{i=2}^{T} y_{t-i}^2\right]^{-1}$	Tables A3.2, or A3.4 asymp.
Model (B): $y_t = \beta_{0c} + \rho_c y_{t-1} + u_t$			
$H_0: \rho_c = 1$	Simple Dickey-Fuller for an arbitrary y_0, and u_t being white noise	$t_{\rho c} = \frac{(\hat{\rho}_c - 1)}{\hat{se}\,(\hat{\rho}_c)}$	Tables A3.2, A3.3 or A3.4
	Normalized bias test for an arbitrary y_0, and u_t being white noise	$T(\hat{\rho}_c - 1)$	Tables A3.2 or A3.4
	Augmented Dickey-Fuller tests for an arbitrary y_0, and u_t being some *ARMA(p,q)*	Based on: $\Delta y_t = \beta_{0c} + (\rho_c - 1) y_{t-1} + \sum_{i=1}^{l} \alpha_i \Delta y_{t-i} + v_t$	
		$t_{\rho c} = \frac{(\hat{\rho}_c - 1)}{\hat{se}\,(\hat{\rho}_c)}$	Tables A3.1 A3.3 or A3.4 asymp
	Perron-Phillips t-test for an arbitrary y_0, and u_t being some *ARMA(p,q)*	$Z(t_{\rho c}) = \left(\frac{S_u}{S_{Tl}}\right) t_{\rho c} - \frac{1}{2}(S_{Tl}^2 - S_u^2)\left[S_{Tl}\left(T^{-2}\sum_{i=2}^{T}(y_{t-i} - \bar{y}_{-1})^2\right)^{1/2}\right]^{-1}$	Tables A3.1, A3.3 or A3.4
	Perron-Phillips normalized bias test for an arbitrary y_0, and u_t being some *ARMA(p,q)*	$Z(\hat{\rho}_c) = T(\hat{\rho}_c - 1) - \frac{1}{2}\left(S_{Tl}^2 - S_u^2\right)\left[T^{-2}\sum_{i=2}^{T}(y_{t-i} - \bar{y}_{-1})^2\right]^{-1}$	Tables A3.2, or A3.4 asymp.

Table 3.11.- Test statistics of unit roots and of other hypotheses in unit roots regressions.[1]

Null hypothesis	Comment	Test statistics	Critical values
$\beta_{0c} = 0$	Dickey-Fuller t-test	$t_{0c} = \hat{\beta}_{0c} / \hat{se}\,(\hat{\beta}_{0c})$	Table A3.5
	Perron-Phillips t-test	$Z(t_{0c}) = \left(\frac{S_u}{S_{Tl}}\right) t_{0c}$	Table A3.5
		$+ (2S_{Tl})^{-1}(S_{Tl}^2 - S_u^2)\left[T^{-2}\sum_{t=2}^{T}(y_{t-1} - \bar{y}_{-1})^2\right]^{-1/2}\left(T^{-2}\sum_{t=2}^{T} y_{t-1}^2\right)^{-1/2}\left(T^{-3/2}\sum_{t=2}^{T} y_{t-1}\right)$	
$\beta_{0c} = 0,\ \rho_c = 1$	F-type test	$\Phi_1 = (2S_u^2)^{-1}\left[(T-1)S_0^2 - (T-3)S_u^2\right]$	Table A3.6
	Perron-Phillips corrected F-type test	$Z(\Phi_1) = \left(\frac{S_u^2}{S_{Tl}^2}\right)\Phi_1 - (2S_{Tl}^2)^{-1}(S_{Tl}^2 - S_u^2)$ $\cdot\left\{T(\hat{\rho}_b - 1) - \frac{1}{4}\left(S_{Tl}^2 - S_u^2\right)\left[T^{-2}\sum_{t=2}^{T}(y_{t-i} - \bar{y}_{-1})^2\right]^{-1}\right\}$	Table A3.6
Model (C): $y_t = \beta_{0ct} + \beta_{1ct}\ t + \rho_{ct} y_{t-1} + u_t$ or $y_t = \beta_{0ct} + \beta_{1ct}\ (t - T/2) + \rho_{ct} y_{t-1} + u_t$			
$\rho_{ct} = 1$	Simple Dickey-Fuller for an arbitrary y_0, and u_t being white noise	$t_{\rho ct} = \frac{(\hat{\rho}_{ct} - 1)}{\hat{se}\,(\hat{\rho}_{ct})}$	Tables A3.1 A3.3 or A3.4
	Normalized bias test for an arbitrary y_0, and u_t being white noise	$T(\hat{\rho}_{ct} - 1)$	Tables A3.2 or A3.4
	Augmented Dickey-Fuller tests for an arbitrary y_0, an u_t being some *ARMA(p,q)*	Based on: $\Delta y_t = \beta_{0ct} + \beta_{1ct}\ t + (\rho_{ct} - 1)\ y_{t-1} + \sum_{i=1}^{l} \alpha_i\ \Delta y_{t-i} + v_t$	
		$t_{\rho ct} = \frac{(\hat{\rho}_{ct} - 1)}{\hat{se}\,(\hat{\rho}_{ct})}$	Tables A3.1 A3.3, or A3.4 asym.
	Perron-Phillips t-test for an arbitrary y_0, and u_t being some *ARMA(p,q)*	$Z(t_{\rho ct}) = \left(\frac{S_u}{S_{Tl}}\right) t_{\rho ct} - \left(\frac{T^3}{4\sqrt{3}}\right)\left(D_x^{-1/2} S_{Tl}^{-1}\right)(S_{Tl}^2 - S_u^2)$	Tables A3.1 A3.3, or A3.4 asym.
	Perron-Phillips normalized bias test for an arbitrary y_0, and u_t being some *ARMA(p,q)*	$Z(\hat{\rho}_{ct}) = T(\hat{\rho}_{ct} - 1) - \left(\frac{T^6}{24}\right)\left(D_x^{-1}\right)\left(S_{Tl}^2 - S_u^2\right)$	Tables A3.2 or A3.4 asym.

Table 3.11.- Test statistics of unit roots and of other hypotheses in unit roots regressions.[1]

Null hypothesis	Comment	Test statistics	Critical values
$\beta_{0c\tau} = 0$	Dickey-Fuller t-test	$t_{0c\tau} = \hat{\beta}_{0c\tau} / \hat{se}\,(\hat{\beta}_{0c\tau})$	Table A3.5
	Perron-Phillips t-test	$Z(t_{0c\tau}) = \left(\frac{S_u}{S_{Tl}}\right) t_{0c\tau}$ $+\, T^3\,(24\,D_x^{1/2} G_x\, S_{Tl})^{-1} (S_{Tl}^2 - S_u^2) \left(T^{-3/2} \sum_{t=2}^{T} y_{t-1}\right)$	Table A3.5
$\beta_{1c\tau} = 0$	Dickey-Fuller t-test	$t_{1c\tau} = \hat{\beta}_{1c\tau} / \hat{se}\,(\hat{\beta}_{1c\tau})$	Table A3.5
	Perron-Phillips t-test	$Z(t_{1c\tau}) = \left(\frac{S_u}{S_{Tl}}\right) t_{1c\tau}$ $-\left\{ T^3\,(2D_x^{1/2}\, S_{Tl})^{-1}\,(S_{Tl}^2 - S_u^2) \left[T^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y}_{-1})^2 \right]^{-1/2} \right\}$ $\times \left\{ \frac{1}{2} \left[T^{-3/2} + T^{-5/2} \right] \sum_{t=2}^{T} y_{t-1} - \sum_{t=2}^{T} t\, y_{t-1} \right\}$	Table A3.5
$\beta_{0c\tau} = 0$, $\beta_{1c\tau} = 0$, $\rho_{c\tau} = 1$	F-type test	$\Phi_2 = (3S_u^2)^{-1} \left[(T-1) S_0^2 - (T-4) S_u^2 \right]$	Table A3.6
	Perron-Phillips corrected F-type test	$Z(\Phi_2) = \left(\frac{S_u^2}{S_{Tl}^2}\right) \Phi_2$ $-\, (3S_{Tl}^2)^{-1} (S_{Tl}^2 - S_u^2) \cdot \left\{ T(\hat{\rho}_{c\tau} - 1) - T^6\,(48\,D_x)^{-1}\,(S_{Tl}^2 - S_u^2) \right\}$	Table A3.6
$\beta_{1c\tau} = 0$, $\rho_{c\tau} = 1$	F-type test	$\Phi_3 = (2S_u^2)^{-1} \left\{ (T-1) S_0^2\, (\bar{y}_0^2 - \bar{y}_1^2) - (T-4) S_u^2 \right\}$	Table A3.6
	Perron-Phillips corrected F-type test	$Z(\Phi_3) = \left(\frac{S_u^2}{S_{Tl}^2}\right) \Phi_3$ $-\, (2S_{Tl}^2)^{-1} (S_{Tl}^2 - S_u^2) \cdot \left\{ T(\hat{\rho}_{c\tau} - 1) - T^6\,(48\,D_x)^{-1}\,(S_{Tl}^2 - S_u^2) \right\}$	Table A3.6
Model (D): $y_t = \beta_{0c\tau\tau} + \beta_{1c\tau\tau}\, t + \beta_{2c\tau\tau}\, t^2 + \rho_{c\tau\tau}\, y_{t-1} + u_t$			
$\rho_{c\tau\tau} = 1$	Simple Dickey-Fuller for an arbitrary y_0, and u_t being white noise	$t_{\rho c\tau\tau} = \frac{(\hat{\rho}_{c\tau\tau} - 1)}{\hat{se}\,(\hat{\rho}_{c\tau\tau})}$	Table A3.4
	Normalized bias test for an arbitrary y_0, and u_t being white noise	$T(\hat{\rho}_{c\tau\tau} - 1)$	Table A3.4

Table 3.11.- Test statistics of unit roots and of other hypotheses in unit roots regressions.[1]

Null hypothesis	Comment	Test statistics	Critical values
	Augmented Dickey-Fuller tests for an arbitrary y_0, and u_t being some *ARMA(p,q)*	Based on: $\Delta y_t = \beta_{0c\tau\tau} + \beta_{1c\tau\tau}\, t + \beta_{2c\tau\tau}\, t^2 + (\rho_{c\tau\tau} - 1)\, y_{t-1} + \sum_{i=1}^{l} \alpha_i\, \Delta y_{t-i} + v_t$	
		$t_{\rho c\tau\tau} = \dfrac{(\hat{\rho}_{c\tau\tau} - 1)}{\hat{se}\,(\hat{\rho}_{c\tau\tau})}$	Table A3.4
where[2]	$S_u^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2;\quad S_{Tl}^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2 + 2T^{-1} \sum_{j=1}^{l} \left[1 - \dfrac{j}{(l+1)}\right] \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}\,;$		
	$D_x = \left[T^2 (T^2 - 1)/12\right] \sum_{t=2}^{T} y_{t-1}^2 - T\left(\sum_{t=2}^{T} t y_{t-1}\right)^2 + T(T+1) \sum_{t=2}^{T} t y_{t-1} \sum_{t=2}^{T} y_{t-1} - \left[\dfrac{T(T+1)(2T+1)}{6}\right]\left(\sum_{t=2}^{T} y_{t-1}\right)^2;$		
	$S_0^2 = (T-1)^{-1} \sum_{t=2}^{T} (y_t - y_{t-1})^2;\quad \bar{y}_0 = (T-1)^{-1} \sum_{t=2}^{T} y_t\,;\quad \bar{y}_1 = (T-1)^{-1} \sum_{t=2}^{T} y_{t-1}\,;$		
	$G_x = \left[T^{-6} D_x + \dfrac{1}{12}\left(T^{-3/2} \sum_{t=2}^{T} y_{t-1}\right)^2\right]^{1/2}.$		

[1] Source: Dickey and Fuller (1981) and Perron (1988)

[2] For each test, the following quantities must be reckoned on the basis of the associated residuals that corresponds to the specific model being entertained.

3.3.3.3 Phillips and Perron tests

Phillips (1987) and Phillips and Perron (1988) have proposed another way to remove the dependency of the asymptotic distribution on the serial correlation of the innovations. They correct nonparametrically the t_ρ-statistic, the normalized bias statistic, the Φ_i-statistic (for $i = 1, 2, 3$), and the other relevant $t_{\beta j}$-statistic (for $j = 0, 1$) computed from the estimates obtained from regressions (3.44). The details of the nonparametric correction of each test are provided in Table 3.11. Essentially, the correction uses a nonparametric estimate of the variance-covariance matrix so that it is free from the influence of serial correlation and of some time-dependent heteroskedasticity.[39] The Phillips-Perron tests are typically denoted by $Z(t_\rho)$, $Z(\rho)$, $Z(\Phi_i)$, and $Z(t_{\beta j})$. They follow the same asymptotic distribution as their non-corrected counterpart would have if there were no serial correlation.

---

[39] Even though there are many alternatives to compute a consistent estimate of the asymptotic variance-covariance matrix, the Newey-West estimator remains to date the most popular employed in the construction of Phillips-Perron tests.

The construction of the nonparametric estimate of the variance-covariance matrix involves generally some weighting functions that require a lag truncation parameter that must grow at a rate no faster than $T^{1/4}$.[40] Schwert (1989) suggested that the Phillips-Perron tests be reported for at least two different values of l that could be chosen according to the following formulae:

$$l_4 = int[4(T/100)^{1/4}] \quad \text{and} \quad l_{12} = int[12(T/100)^{1/4}], \tag{3.48}$$

where $int[\bullet]$ is the integer function that returns the largest integer that is smaller than or equal to its argument (e.g. $int[3.14] = 3$). Good practice should therefore incorporate a sensitivity check of the results to different values of the lag truncation parameter l.

3.3.3.4 Statistical size of the tests and the nature of the underlying stochastic process

3.3.3.4.1 A comparison of ADF tests and Phillips-Perron tests

Handa and Ma (1989) have conducted Monte Carlo experiments suggesting that the ADF tests and their variants have better small sample properties with non-normal innovations or with heteroskedastic errors. However, with autocorrelated errors or in large samples, the Phillips-Perron tests perform better.

3.3.3.4.2 Unit root tests and negative MA component in the innovation

One important complication arises when performing unit root tests on time series whose first-difference exhibits negative moving average serial correlation. Schwert's (1987a,b, 1989) Monte-Carlo simulations have shown that in small samples (even with 1000 observations), the Phillips and Perron tests and the normalized bias test will reject too often the unit root hypothesis when compared to tabulated critical values. These biases in the tabulated critical values become also more important as the $MA(1)$ coefficient tends to -1 in an $IMA(1,1)$ series. Indeed, considering an $IMA(1,1)$ process $y_t - y_{t-1} = \varepsilon_t + \theta\, \varepsilon_{t-1}$ with white noise innovations, when θ tends to -1, the level of the variable y_t is near stationary. In finite samples, y_t may well look more like a white noise than like an integrated process.

However, Schwert (1987a, 1989) finds that the critical values for the ADF statistic appeared to be less sensitive to the nature of the innovation process, particularly if a high AR approximation, say with l_{12} defined by equation (3.48), is used in the ADF regression equation, following Said and Dickey (1984). In practice, it is therefore advisable to look at the appearance of the sample ordinary and partial correlogramms of the first-difference of a series, and even to estimate an *IMA*(1,1) for the series under consideration. This could warn us for cases where the Phillips-Perron tests and the normalized bias test would lead to incorrect inference about the presence of unit roots.

---

[40] Despite the existence of a lag truncation parameter in the Newey-West estimator, the choice of l is not motivated on the same basis as the lag truncation parameter in an augmented Dickey-Fuller regression.

If negative MA terms seem to characterize the series under consideration, there are several courses of action. First, one could decide not to rely on Phillips-Perron tests or the normalized bias test.[41] Second, one could compare the calculated values of these statistics with critical values computed by Schwert (1987a) for an IMA(*1,1*). These critical values are reproduced in Tables A3.7 and A3.8. Thirdly, Said and Dickey's (1985) Monte Carlo experiments suggested that the ADF *t*-test, based on the estimation of equations similar to (3.47) but augmented to estimate also the MA(*k*) terms, asymptotically follows the Dickey-Fuller distribution. However, Schwert's (1987a) Monte Carlo experiments show that this is true at least from a 5% significance level, but that it is not so at the 1% significance level. His results suggest a rather flat-tail distribution for this test. Finally, Hall (1989), and Pantula and Hall (1991) recommend another procedure that does not require the estimation of the MA parameters. One has to estimate the augmented regressions (3.47) with an instrumental variable technique rather than with OLS before applying the ADF test. To test an ARIMA(*l,1,k*) against an ARIMA(*l+1,0,k*), $y_{t-(k+1)}$ can be used as the level instrumental variable associated to y_{t-1}, and the vector $[\Delta y_{t-(k+1)}, \Delta y_{t-(k+1)-1}, \ldots, \Delta y_{t-l-(k+1)+1}]'$ can be used as instruments for $[\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-l}]'$.[42, 43]

---

[41] However, Maekawa's (1994) Monte Carlo experiments with an IMA(*1,1*) suggest that if one prewhitens the series y_t by regressing it on, say, *k* lags of itself, the $Z(\rho)$ statistic that is applied on the residual of the prewhitening regression has a distribution close to the Dickey-Fuller distribution in large sample. This prewhitened unit root test seems to perform generally well and is preferable to Phillips-Perron test, except when θ is close to zero.

[42] Technically, this procedure, as well as Said and Dickey's (1985), requires that *l*, and an upperbound for *k* be known, since these numbers determine the specification of the regression equation on which the test is based. The value of *l* determines the number of lag first-differences to include, while *k* determines the instruments to be used. In practice, one could choose *k* by examining the sample autocorrelation function of the first differenced data or of the residuals of the ADF equation, which should have non-zero elements up to lag *k*. However, this means that the selection of *l* and *k* may presuppose the existence of a unit root, which may affect adversely the power of the unit root test. An alternative has been proposed by Hall (1992) who developped a residual based test statistic to estimate *l* and *k*.

[43] Pantula and Hall (1991) found that overspecification of *l* and/or *k* may not generally affect too much the performance of the IV based ADF t_ρ-test. However, underspecification will certainly lead to an inconsistent statistic with the IV procedure. Moreover, the erroneous inclusion of an intercept and/or a linear trend when there are negative MA innovations will cause rejections of the unit root more often than the adopted significance level. Said and Dickey's (1984) basic ADF test

3.3.3.5 Power of unit root tests

In practice, it is not very easy to distinguish between a unit root process (with $\rho = 1$) and a near unit root process (with $\rho = 0.95, 0.98$ or 0.99). With a near-unit root process, the series y_t eventually returns to its mean even though it may take quite some time, i.e. there is a lot of persistence. Unit root tests will generally have little power to discriminate between an integrated process and a near-integrated process in finite samples. In fact, Campbell and Perron (1991) provide simulation results that suggest that "near-stationary unit root DGPs are better forecast using stationary forecasts models, while near-integrated stationary DGPs are better forecast using integrated forecasting models." (p. 162). Another way to look at the power problems of unit root tests is that one typically has little information to infer the cyclical behaviour of the first-difference of a series at the zero frequency (i.e. an infinitely long period).

Phillips and Ouliaris (1990) have shown that the normalized bias tests and its variants are theoretically more powerful than the t_ρ- or the Φ_i-tests. This is essentially because of the faster rate of convergence of the normalized bias test at rate T^{-1} rather than $T^{-1/2}$.

As shown by Perron and Shiller (1987), the power of unit root tests also depends much more on having a long span of data rather than the frequency of the data available. Hence, one would not generally gain more power in unit root tests on output by working with 288 monthly observations rather than with 24 annual observations.

In some instances, we can conduct both joint hypotheses tests and single hypothesis unit root tests. However, DeJong *et al.* (1992) find that the normalized bias statistic and the t_ρ-statistic and their variants both have more power than the Φ-statistic and its variant in practice.

3.3.3.6 Unit root tests and structural breaks

Conclusions drawn from unit root tests may well be sensitive to structural breaks in the underlying stochastic processes. Perron (1989) showed how a break in a deterministic trend at a particular date (say at the start of the Great Depression or at the outset of the first OPEC oil

---

based on a high-order AR approximation of the MA part of the process may be less sensitive to this problem. Finally, Hall (1992) finds that there may be some combinations of parameter values that make the Dickey-Fuller distribution a deficient approximation of the distribution of the IV based ADF t_ρ-test.

shock) may bias the integration tests to support statistically the unit root hypothesis in real output. Perron (1990) makes a similar point for series with changing mean. Christiano (1988) has criticized Perron's (1989) methodology of suffering from pre-testing bias in the sense that the evidence of a broken trend-stationary process was gathered from a regression where some dummy variables were defined at a date chosen after having looked at the data. Reassessing Perron's evidence to correct the critical values of the test for pre-testing bias, Christiano (1992) does not reject the unit root hypothesis in post-war quarterly real GNP.[44] Other tests of a unit root against trend stationarity with break in the trend and estimated breakpoint, have been developped by Banerjee, Lumsdaine and Stock (1992), and Zivot and Andrews (1992). They also fail to reject the unit root hypothesis for post-war quaterly real GNP.[45]

There is certainly room for different interpretation when attempting to discriminate between an integrated process with drift and a broken trend stationary process. One may argue that what appears to some to be a break in a deterministic trend (which as a permanent effect) may be only one big positive or negative realization of the innovation of the same underlying stochastic process. Indeed, as illustrated in panel B of Table 3.10, a random walk with drift is essentially analogue to a deterministic trend that breaks at each period. Also, an observed time series could be drawn from a random walk with drift over the first portion of the sample while the second portion would be generated by another random walk with a different drift. In this case, it might happen that the difference-stationary hypothesis be rejected for the variable over the whole sample.

Nonetheless, the possibility raised by Perron (1989, 1990) that evidence favouring unit root tests may be resulting from breaks in the underlying DGP justifies that one be careful when working with a long period of data. As suggested earlier, an examination of recursive estimates of the sample autocorrelation coefficients, say up to the 6th order, can be instructive.

3.3.3.7 Unit root tests and officially seasonally adjusted series

Davidson and MacKinnon (1993) argue that unit root tests applied on time series, that have been seasonally adjusted with linear filters or official statistical agencies procedures that can often be approximated by some linear filters, will tend to provide evidence of integration even when the underlying unadjusted DGP is stationary. Ghysels and Perron (1993) conducted Monte Carlo experiments that supported this analysis. This suggests again that it is preferable when possible to work with unadjusted raw data for which one can identify some seasonal *ARIMA* process, if necessary. In either case, Ghysels and Perron (1993) recommend that ADF regressions be estimated with a maximum autoregression of order l_{max} greater than the seasonal lag, when applying Campbell and Perron's (1991) recursive data dependent determination of l.

---

[44] Other recent papers have also reported mixed evidence about the presence or not of integration in macroeconomic time-series. See Kim (1990) and Rudebusch (1990).

[45] Readers interested in applying these tests are refer to the articles mentionned above as well as to Perron and Vogelsang (1992).

3.3.4 Strategy of unit root testing and applications

Having discussed many of the existing unit root tests one can employ, as well as caveats one must be aware of, let us now summarize the steps that can be followed. Perron (1988), and Campbell and Perron (1991) suggested a sequential testing procedure starting from the general-to-specific representation. This would amount to begin one's quest with model (3.47d), and then follow a testing down approach. Specification (3.47d) has mostly been ignored in the applied literature and the critical values for a joint test of integration and zero coefficient on t^2 has not been computed. Therefore, equation (3.47c) is generally the starting model considered in practice.

The following is a set of suggested rules to be applied sequentially.

1. Visual and preliminary inspection of the data

   (i) Examine a plot of the series and attempt to identify whether the series exhibits a drift.

   (ii) Examine the sample autocorrelation function of the series and look for lengthy dying of the autocorrelations.

   (iii) Compute and examine also recursive estimates of the sample autocorrelations to check for visible shift or break in the underlying stochastic process of the series.
2. Estimate an *IMA*(1,1) representation for the series. If there is a negative MA coefficient < -0.2, beware of Schwert's critical values and of the limitations of Phillips-Perron tests. You may also want to consider Pantula and Hall (1991) IV based ADF test.
3. Compute the Cochrane variance ratio test.

4. Estimate model (3.47c):

$$\Delta y_t = \beta_{0ct} + \beta_{1ct}\, t + (\rho_{ct} - 1)\, y_{t-1} + \sum_{i=1}^{l} \alpha_i\, \Delta y_{t-i} + v_t \ .$$

(i) Test down the lag truncation parameter l. If $l = 0$, you can compute also the normalized bias test where applicable.

(ii) Compute the statistics: t_ρ to test H_0: $\rho = 1$, i.e. a unit root, and Φ_3 to test H_0: $\rho = 1$, $\beta_1 = 0$, i.e. a random walk with drift against H_1: the trend-stationary hypothesis. If either test rejects the null, the series should be considered trend-stationary; the testing may end for this step.[46] If not, there is little evidence so far against a unit root in the time-series or against a drift.

(iii) Compute the Φ_2 statistics to test H_0: $\rho = 1$, $\beta_1 = 0$, $\beta_0 = 0$. If the null hypothesis is not rejected, there is not enough evidence supporting a drift. Because the t_ρ may be more powerful in specification (3.47b), we can proceed with step 5. If it is rejected, this suggests a difference-stationary process with drift; the testing may end for this step.

5. Estimate model (3.47b):

$$\Delta y_t = \beta_{0c} + (\rho_c - 1)\, y_{t-1} + \sum_{i=1}^{l} \alpha_i\, \Delta y_{t-i} + v_t \ .$$

(i) Test down the lag truncation parameter l. If $l = 0$, you can compute also the normalized bias test where applicable.

(ii) Compute the statistics: t_ρ to test H_0: $\rho = 1$, i.e. a unit root, and Φ_1 to test H_0: $\rho = 1$, $\beta_0 = 0$, i.e. a random walk without drift against H_1: the mean-stationary hypothesis. If either test rejects the null, the series should be considered mean-stationary; the testing may end for this step. If not, there is little evidence so far against a unit root in the time-series.

6. Estimate the model (3.44c):

$$\Delta y_t = \beta_{0ct} + \beta_{1ct}\, t + (\rho_{ct} - 1)\, y_{t-1} + u_t \ .$$

(i) Compute the statistics: $Z(t_\rho)$ and $Z(\rho)$ to test H_0: $\rho = 1$, i.e. a unit root, and $Z(\Phi_3)$ to test H_0: $\rho = 1$, $\beta_1 = 0$, i.e. a random walk with drift against H_1:the trend-stationary hypothesis. If either test rejects the null, the series should be considered trend-stationary; the testing may end for this step. If not, there is little evidence so far against a unit root in the time-series or against a drift.

---

[46] Notice that conflicting results can arise in finite samples from different tests.

(ii) Compute the $Z(\Phi_2)$ statistics to test H_0: $\rho = 1$, $\beta_1 = 0$, $\beta_0 = 0$. If the null hypothesis is not rejected, there is not enough evidence supporting a drift. Because the $Z(t_\rho)$ may be more powerful in specification (3.44b), we can proceed with step 5. If it is rejected, this suggests a difference-stationary process with drift; the testing may end for this step.

7. Estimate the model (3.44b):

$$\Delta y_t = \beta_{0c} + (\rho_c - 1)\, y_{t-1} + u_t \quad .$$

Compute the statistics: $Z(t_\rho)$ and $Z(\rho)$ to test H_0: $\rho = 1$, i.e. a unit root, and $Z(\Phi_1)$ to test H_0: $\rho = 1$, $\beta_0 = 0$, i.e. a random walk without drift against H_1: the mean-stationary hypothesis. If either test rejects the null, the series should be considered mean-stationary; the testing may end for this step. If not, there is little evidence against a unit root in the time-series.

At first sight, applying all the above steps may seem somewhat tedious, but it is fairly easy to write a program (e.g. in Eviews, in GAUSS, or in RATS) that automates the procedure and that report all the relevant statistics. If such a program is not easily accessible, one can skip steps 6 and 7 involving non-parametrically corrected unit root tests.

3.3.5 Cochrane variance ratios and unit root tests for the log Canadian exchange rate and its first difference.

Table 3.12 shows the evolution of the Cochrane variance ratio with respect to different values of k, for the logarithm of the Canada-US exchange rate series, as well as for its first-difference. The point estimate of the ratio for the original series tends to converge to a positive constant, which suggests that $\log(e_t) \sim I(1)$, while that of $\Delta\log(e_t)$ converges quickly to a value not different from zero. This supports the view that $\Delta\log(e_t) \sim I(0)$.

Notice that as k increases the confidence interval widens around the estimate of V^k. Relative to their standard-errors, the variance ratios of the two series are about the same, which reflects how little information one has ultimately about the very long run. Nonetheless, the variance ratio series suggest the order of integration of the Canadian exchange rate series.

Table 3.13 presents augmented Dickey-Fuller tests for various specifications. For each equation, the lag length l was chosen using Campbell and Perron's (1991) recursive procedure, with a 10% significance level. For the log-level series, neither the deterministic linear trend, nor the intercept were significant. The *ADF* test did not reject the unit root hypothesis even at a 10% significance level. When performing the test on the log- first-differenced exchange rate, the final specification that was adopted had no constant and no linear trend. The *ADF*-statistic rejected the unit root hypothesis at the 1% level.

We have also conducted unit root tests that we show in Table 3.14. The lag lengths, used to construct the Newey-West non-parametric correction for serial correlation and heterogeneity, were chosen according to formulae (3.48), following Schwert (1989). Hence, $l_4 = int[4\,(241/100)^{1/4}] = 4$, and $l_{12} = int[12\,(241/100)^{1/4}] = 14$. None of the Perron-Phillips tests that were applied on $log(e_t)$ could reject the unit root hypothesis at the 10% significance level. In particular, the joint test of a unit root and no time varying drift, and the joint test of a unit root and zero intercept could not be rejected either. Finally, all the Phillips-Perron tests that were applied on the first-difference of $log(e_t)$ rejected clearly the unit root hypothesis at least at the 1% significance level.

Hence, the formal analysis of this section confirms the first-order integration of the log-exchange rate series, previously postulated in section 3.2.

Table 3.12- Cochrane variance ratios for the Canada-US exchange rate (in log) and its first difference (1973:01-1993:01 sample; 241 observations). (Standard-error in parentheses).

Series	V^{15}	V^{30}	V^{40}	V^{60}	V^{100}	V^{120}
$\log(e_t)$	1.7461 (0.5091)	2.0612 (0.8366)	2.1545 (1.0056)	2.0775 (1.1828)	1.2750 (0.9341)	1.1875 (0.9522)
$\Delta\log(e_t)$	0.0760 (0.0222)	0.0433 (0.0176)	0.0289 (0.0135)	0.0223 (0.0127)	0.0129 (0.0095)	0.0110 (0.0089)

Table 3.13 - Augmented Dickey-Fuller tests for the Canadian exchange rate (in log) and its first difference. Sample: 1973:01-1993:01.[1]

Series	Equation (3.47c)						Equation (3.47b)					Equation (3.47a)			
	Lag length l	Significance level on			Estimate		Lag length l	Significance level on		Estimate		Lag length l	Sign. level	Estimate	
		l	β_{0ct}	β_{1ct}	ρ_{ct}	$t_{\rho ct}$		l	β_{0c}	ρ_c	$t_{\rho c}$		l	ρ_n	$t_{\rho n}$
$\log(e_t)$	10	0.001	>.20	>.20	0.9819	-2.05	12	0.071	>.20	0.9882	-1.79	12	0.061	0.9997	-0.08
$\Delta\log(e_t)$	12	0.099	>.20	>.20	0.3538	-3.50 **	11	0.057	>.20	0.4713	-3.06 **	11	0.060	0.5084	-2.91 *

[1] Rejection level of H_0: $\rho = 1$: * at 1%; ** at 5%; *** at 10%.

Table 3.14 - Perron-Phillips unit root tests for the Canadian exchange rate (in log) and its first difference. Sample: 1973:01-1993:01.[1]

Series	Lag length	$Z(\hat{\rho}_{ct})$	$Z(t_{\rho c})$	$Z(\Phi_3)$	$Z(\Phi_2)$	$Z(\hat{\rho}_c)$	$Z(t_{\rho c})$	$Z(\Phi_1)$
$\log(e_t)$	$l_4 = 4$	-2.893	-1.209	0.952	1.494	-2.501	-1.362	2.119
	$l_{12} = 14$	-3.944	-1.409	1.156	1.398	-2.874	-1.416	1.978
$\Delta\log(e_t)$	$l_4 = 4$	-200.77 *	-12.95 *	83.88 *	56.47 *	-13.65 *	-253.57 *	93.58 *
	$l_{12} = 14$	-232.04 *	-13.33 *	88.95 *	59.74 *	-296.18 *	-14.28 *	102.29 *

[1] Rejection level of H_0: * at 1%; ** at 5%; *** at 10%.

3.3.6 Unit roots in macroeconometrics

Having reviewed some of the most popular unit root tests employed in the literature, we are left with the issue regarding to which extent evidence of unit roots is important from a theoretical and an applied macroeconomic point of view. Christiano and Eichenbaum (1990) have in a sense best summarized this point in what may be the shortest abstract ever written: "Do we Know? Do we care? -- Perhaps. Maybe not." On the first ground, the low power of unit root tests makes it very difficult to distinguish between integrated series and nearly integrated series, especially with a finite size sample. Furthermore, the issue of persistence that mostly matters in macroeconomics does not require the integration type of persistence. From an economic agent's intertemporal and forward-looking perspective, the distinction between permanent versus temporary changes in a macroeconomic variable need not correspond to the integration of this variable. The following example reported in Maddala (1992) illustrates this point.

Suppose that the variable y_t has the following DGP,

$$y_t = \rho\, y_{t-1} + \varepsilon_t \tag{3.49}$$

where ε_t is white noise. If $|\rho| < 1$, then the process is stationary and a one time shock of size S at date $t+s$ will impact on future values of y_t as follows: $\Delta y_{t+s} = S$, $\Delta y_{t+s+1} = \rho\, S$, $\Delta y_{t+s+2} = \rho^2 S$, $\Delta y_{t+s+3} = \rho^3 S$, etc. However, if the process is integrated, i.e. $\rho = 1$, then $\Delta y_{t+s} = S$, $\Delta y_{t+s+1} = S$, $\Delta y_{t+s+2} = S$, $\Delta y_{t+s+3} = S$, etc. Now, if economic agents care about the present value of future streams of y_t and use a constant discount rate r, the change in this present value as of date $t+s$ is given by

$$\begin{aligned} \Delta PV(y_{t+s}) &= \Delta y_{t+s} + \frac{\Delta y_{t+s+1}}{1+r} + \frac{\Delta y_{t+s+2}}{(1+r)^2} + \frac{\Delta y_{t+s+3}}{(1+r)^3} + \cdots \\ &= S + \frac{\rho S}{1+r} + \frac{\rho^2 S}{(1+r)^2} + \frac{\rho^3 S}{(1+r)^3} + \cdots \\ &= \frac{S}{[1-\rho/(1+r)]}\,. \end{aligned} \tag{3.50}$$

While it is true that whether ρ is equal to or different from one determines whether the effect of an innovation on the variable y_t is permanent or transitory, if what is of concern depends on the change in the present value of y_t, then

$$\Delta PV(y_{t+s}) = \begin{cases} \frac{S}{1-\rho} < \infty & \text{if } \rho < 1 \text{ and } r = 0 \\ \infty & \text{if } \rho = 1 \text{ and } r = 0 \\ \frac{(1+r)\,S}{r} < \infty & \text{if } \rho = 1 \text{ and } r > 0 \\ < \infty & \text{if } \rho < 1 + r \end{cases} . \qquad (3.51)$$

Hence, the existence of a unit root will not generally lead to explosive changes in the present value of current and future streams of a variable. Most likely, the macroeconomic environment in which economic agents evolve is affected by many transitory and permanent disturbances, not solely by a single permanent shock as we could be concluding with too much hastiness from evidence supporting a unit root in real GDP (or GNP).

Evidence favouring the unit root hypothesis has also been sometimes referred as allowing one to discriminate between macroeconomic theories of business cycles. For instance, a pure simple version of the real business cycle theory implies that all movements in real GNP are due to real factors, while a Keynesian theory would rely mostly upon monetary factors to explain business cycles. Since the consensus in macroeconomics tends to acknowledge that money is neutral in the long-run, it is tempting to conclude that evidence of a unit root in real GNP supports rather the idea that real factors underlie mostly the business cycles. This was first mentioned in Nelson and Plosser (1982) and was argued recently in Maddala (1992) to explain why unit roots may matter from a theoretical macroeconomic point of view. Unfortunately, things are not so simple and an integrated real GNP does not necessarily lead to the rejection of one theory. Indeed, Ambler and Phaneuf (1989) and West (1988b) have constructed theoretical representations of artificial economies with nominal wage contracts and technology shocks for which real output may exhibit a near-unit root, or a unit root even though the state of the technology was not integrated.

Hence if it is not so theoretically helpful to gather evidence of unit roots in macroeconomic time series and if the tests we perform are not necessarily so powerful, does this mean that we should stop worrying about unit roots and go back to the old days when this issue was ignored ? The answer is a resounding no for the following reasons.

First, the lengthy persistent effects of innovations on many macroeconomic variables (e.g. GNP) must be taken as an atheoretical characteristic of the series. It must be reproduced in any relevant models of business cycles, even though it does not discard all but one class of models. Second, non-spurious relationships between integrated series that can be found in the data may sometimes be interpreted as long-run macroeconomic equilibria that are worth checking in the data. These are called cointegrating relationships and will be the object of many sections in the next chapter. In the same vein, it will be important to avoid the estimation of spurious regressions between independent $I(1)$ variables or the estimation of unbalanced regressions (e.g. using an $I(1)$ regressand on only $I(0)$ regressors).

Other reasons are econometric in nature. The need to characterize data has trend-stationary (TS) or has difference-stationary (DS) pertains when one wishes to remove the trend in time-series. As shown by Nelson and Kang (1981), the wrong differencing of TS data or of deterministic detrending of a DS series will result in spurious autocorrelation in the transformed series. Finally, as we will discuss in further details in the next chapter, the least-squares estimator (and its variants) of regression coefficients associated with integrated series are generally not asymptotically normally distributed so that the application of standard inference too will be distorted in such instances.[47]

3.4 Summary.

Univariate time-series methods consist in useful tools to characterize the evolution of macroeconomic variables over time, based on the data at hands. We have discussed how the underlying deterministic or stochastic nature of a trend in a series was crucial to model stochastic processes. This will also have important implications that we will need to deal with in the next chapter. Various unit root tests were discussed in terms of both their construction and their statistical properties.[48] Once the appropriate order of integration of a series is established, a stationary series can be constructed by differencing enough times the original level series. Then, some stationary stochastic process of the differenced series must be identified before estimation. Finally, the estimated model needs to be tested for misspecification.

---

[47] Campbell and Perron (1991) have argued that unit root tests can be helpful in suggesting the imposition of restrictions on near-integrated series to obtain estimates with lower mean squared errors. Furthermore, they suggested that integrated asymptotic econometric theory may be preferable to the asymptotically standard normal distribution theory, for nearly integrated time series in finite samples. Ideally, it would however be even preferable to calculate the finite sample distributions with extensive Monte-Carlo experiments.

The econometrics of univariate time series involves many concepts that are nonetheless integrated in a fairly unified framework. What matters is to proceed methodically in the same vein as suggested by Hendry's methodology.

1. First, do not neglect the visual inspection of the series itself, in levels and in first-differences. Examine also the sample ordinary and partial autocorrelation functions of the series in levels and in first-differences. (You may need to difference more than once, if the order of integration is greater than one).
2. If there is suspicion that the series may be non-stationary, perform unit root tests on it. You may also compute Cochrane's variance ratio.
3. Once, the order of integration is established, construct a stationary series by appropriate differenciation of the original series.
4. From examining the sample ordinary and partial autocorrelation functions of the stationary series, along with Hannan and Rissanen procedure, attempt to identify the orders of possibly appropriate *ARMA* models, with seasonal components if necessary.
5. Submit any non-discarded models to diagnostic tests especially for serial correlation, for heteroskedasticity tests, for normality, and for stability.

If one is not interested in univariate modelling of some series, but is concerned only with their possible integration, only steps 1 and 2 need to be applied.

---

[48] There are also other unit root tests that have not been discussed in this chapter. A class of tests have been developed to test the null hypothesis of stationarity against the alternative of integration. References to these tests are: Kahn and Ogaki (1990), Park (1990), and Park and Choi (1988). Sims (1988) has also developed a test based on Bayesian posterior odds ratios to discriminate between a unit root and a large but stationary AR coefficient. Finally, Choi and Phillips (1993) proposed frequency domain unit root tests. Their Monte Carlo experiments suggest that these latter tests have stable sizes and good power in finite samples, although their distribution remains sensitive to negative MA serial correlation in the errors.

3.5 *Appendix Tables.*

Table A3.1 - Critical values of the Dickey-Fuller test $t_\rho = (\hat{\rho} - 1)/se(\hat{\rho})$ for $H_0: \rho = 1$ vs. $H_1: \rho < 1$.[1]

Sample size T	Model (A)				Model (B)				Model (C)			
	No constant, no trend				Constant, no trend				Constant & trend			
	Probability of a smaller value				Probability of a smaller value				Probability of a smaller value			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-2.66	-2.26	-1.95	-1.60	-3.75	-3.33	-3.00	-2.63	-4.38	-3.95	-3.60	-3.24
50	-2.62	-2.25	-1.95	-1.61	-3.58	-3.22	-2.93	-2.60	-4.15	-3.80	-3.50	-3.18
100	-2.60	-2.24	-1.95	-1.61	-3.51	-3.17	-2.89	-2.58	-4.04	-3.73	-3.45	-3.15
250	-2.58	-2.23	-1.95	-1.62	-3.46	-3.14	-2.88	-2.57	-3.99	-3.69	-3.43	-3.13
∞	-2.58	-2.23	-1.95	-1.62	-3.43	-3.12	-2.86	-2.57	-3.96	-3.66	-3.41	-3.12

[1] Source: Fuller (1976), Table 8.5.2

Table A3.2 - Critical values of the normalized bias test $T(\hat{\rho} - 1)$ for $H_0: \rho = 1$ vs. $H_1: \rho < 1$.[1]

Sample size T	Model (A)				Model (B)				Model (C)			
	No constant, no trend				Constant, no trend				Constant & trend			
	Probability of a smaller value				Probability of a smaller value				Probability of a smaller value			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-11.9	-9.3	-7.3	-5.3	-17.2	-14.6	-12.5	-10.2	-22.5	-19.9	-17.9	-15.6
50	-12.9	-9.9	-7.7	-5.5	-18.9	-15.7	-13.3	-10.7	-25.7	-22.4	-19.8	-16.8
100	-13.3	-10.2	-7.9	-5.6	-19.8	-16.3	-13.7	-11.0	-27.4	-23.6	-20.7	-17.5
250	-13.6	-10.3	-8.0	-5.7	-20.3	-16.6	-14.0	-11.2	-28.4	-24.4	-21.3	-18.0
500	-13.7	-10.4	-8.0	-5.7	-20.5	-16.8	-14.0	-11.2	-28.9	-24.8	-21.5	-18.1
∞	-13.8	-10.5	-8.1	-5.7	-20.7	-16.9	-14.1	-11.3	-29.5	-25.1	-21.8	-18.3

[1] Source: Fuller (1976), Table 8.5.1

Table A3.3- Parameter values necessary to compute MacKinnon's (1991) critical values of t_ρ.[1]

$H_0: \rho = 1$ vs. $H_1: \rho < 1$

The critical value at the *p*% significance level: $C(p) = \varphi_\infty + \varphi_1 T^{-1} + \varphi_2 T^{-2}$

Model	*p* %	$\hat{\varphi}_\infty$	$\hat{\varphi}_1$	$\hat{\varphi}_2$
(A) No constant, no trend	1%	-2.5658	-1.960	-10.04
	5%	-1.9393	-0.398	0.0
	10%	-1.6156	-0.181	0.0
(B) Constant, no trend	1%	-3.4336	-5.999	-29.25
	5%	-2.8621	-2.738	-8.36
	10%	-2.5671	-1.438	-4.48
(C) Constant & trend	1%	-3.9638	-8.353	-47.44
	5%	-3.4126	-4.036	-17.83
	10%	-3.1279	-2.418	-7.58

[1] Source: MacKinnon (1991)

Table A3.4- Asymptotic critical values based on response surfaces[1]

$H_0: \rho = 1$ vs. $H_1: \rho < 1$

Test statistic	Size of the test			
	0.01	0.025	0.05	0.10
$t_{\rho n}$	-2.56	-2.23	-1.94	-1.62
$t_{\rho c}$	-3.43	-3.12	-2.86	-2.57
$t_{\rho ct}$	-3.96	-3.66	-3.41	-3.13
$t_{\rho ctt}$	-4.37	-4.08	-3.83	-3.55
$T(\hat{\rho}_n - 1)$	-13.7	-10.4	-8.0	-5.7
$T(\hat{\rho}_c - 1)$	-20.6	-16.9	-14.1	-11.2
$T(\hat{\rho}_{ct} - 1)$	-29.4	-25.1	-21.7	-18.2
$T(\hat{\rho}_{ctt} - 1)$	-36.6	-31.8	-28.1	-24.2

[1] Source: Davidson and MacKinnon (1993)

Table A3.5 - Critical values for the t_{0c}, t_{0ct}, and t_{1ct} statistics and their variants[1]

Sample size T	t_{0c} in model (B)				t_{0ct} in model (C)				t_{1ct} in model (C)			
	Size of the test				Size of the test				Size of the test			
	0.10	0.05	0.025	0.01	0.10	0.05	0.025	0.01	0.10	0.05	0.025	0.01
25	2.20	2.61	2.97	3.41	2.77	3.20	3.59	4.05	2.39	2.85	3.25	3.74
50	2.18	2.56	2.89	3.28	2.75	3.14	3.47	3.87	2.38	2.81	3.18	3.60
100	2.17	2.54	2.86	3.22	2.73	3.11	3.42	3.78	2.38	2.79	3.14	3.53
250	2.16	2.53	2.84	3.19	2.73	3.09	3.39	3.74	2.38	2.79	3.12	3.49
500	2.16	2.52	2.83	3.18	2.72	3.08	3.38	3.72	2.38	2.78	3.11	3.48
∞	2.16	2.52	2.83	3.18	2.72	3.08	3.38	3.71	2.38	2.78	3.11	3.46

[1] Source: Dickey and Fuller (1981), Tables I, II and III.

Table A3.6 - Critical values for the Φ_1, Φ_2 and Φ_3 statistics and their variants[1]

Sample size T	Φ_1 in model (B)				Φ_2 in model (C)				Φ_3 in model (C)			
	Size of the test				Size of the test				Size of the test			
	0.10	0.05	0.025	0.01	0.10	0.05	0.025	0.01	0.10	0.05	0.025	0.01
25	4.12	5.18	6.30	7.88	4.67	5.68	6.75	8.21	5.91	7.24	8.65	10.61
50	3.94	4.86	5.80	7.06	4.31	5.13	5.94	7.02	5.61	6.73	7.81	9.31
100	3.86	4.71	4.63	4.61	4.16	4.88	5.59	6.50	5.47	6.49	7.44	8.73
250	3.81	4.63	5.45	6.52	4.07	4.75	5.40	6.22	5.39	6.34	7.25	8.43
500	3.79	4.61	5.41	6.47	4.05	4.71	5.35	6.15	5.36	6.30	7.20	8.24
∞	3.78	4.59	5.38	6.43	4.03	4.68	5.31	6.09	5.34	6.25	7.16	8.27

[1] Source: Dickey and Fuller (1981), Tables IV, V and VI.

Table A3.7 - Schwert's critical values of t- unit root tests when the DGP is an IMA(1,1) with negative serial correlation: $\Delta y_t = \varepsilon_t + \theta\,\varepsilon_{t-1}$.[1, 2]

Sample size	MA parameter	Constant, no trend								Constant and trend							
		$Z(t_{\rho\mu})$ and l_4		$Z(t_{\rho\mu})$ and l_{12}		ADF $t_{\rho\mu}$ with an estimated ARMA(1,1)		ADF $t_{\rho\mu}$ with l_{12}		$Z(t_{\rho\tau})$ and l_4		$Z(t_{\rho\tau})$ and l_{12}		ADF $t_{\rho\tau}$ with an estimated ARMA(1,1)		ADF $t_{\rho\tau}$ with l_{12}	
T	θ	Size of the test		Size of the test		Size of the test		Size of the test		Size of the test		Size of the test		Size of the test		Size of the test	
		0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
25	0.0	-3.74	-3.01	-3.94	-3.05	-6.20	-2.71	-3.64	-2.87	-4.39	-3.62	-4.69	-3.57	-7.27	-3.31	-4.31	-3.48
	-0.5	-5.73	-4.76	-6.11	-4.89	-7.82	-3.84	-3.63	-2.85	-6.55	-5.55	-7.81	-6.01	-9.17	-3.83	-4.23	-3.40
	-0.8	-7.44	-6.35	-8.54	-6.80	-7.25	-4.05	-3.57	-2.85	-7.96	-6.88	-10.47	-8.34	-8.42	-3.70	-4.25	-3.40
50	0.0	-3.60	-2.97	-3.62	-2.92	-3.19	-2.56	-3.45	-2.80	-4.17	-3.55	-4.06	-3.39	-3.48	-2.78	-4.03	-3.37
	-0.5	-6.09	-5.14	-6.34	-5.46	-5.72	-2.65	-3.41	-2.78	-6.94	-6.06	-7.02	-6.19	-5.47	-2.93	-3.94	-3.32
	-0.8	-8.64	-7.62	-8.85	-7.79	-7.74	-4.18	-3.55	-2.90	-9.28	-8.29	-10.27	-8.72	-9.06	-4.11	-4.11	-3.46
100	0.0	-3.57	-2.95	-3.58	-2.95	-3.43	-2.76	-3.48	-2.82	-4.07	-3.53	-3.97	-3.45	-3.61	-3.06	-3.95	-3.34
	-0.5	-6.47	-5.34	-7.16	-6.07	-3.22	-2.54	-3.55	-2.80	-7.48	-6.55	-7.95	-7.10	-3.20	-2.55	-3.91	-3.36
	-0.8	-10.25	-9.36	-10.70	-9.93	-6.76	-2.96	-3.42	-2.93	-11.09	-10.13	-11.28	-10.39	-6.40	-3.37	-4.16	-3.49
250	0.0	-3.47	-2.89	-3.51	-2.94	-4.54	-3.03	-3.41	-2.82	-4.09	-3.52	-4.08	-3.53	-3.82	-3.27	-3.95	-3.40
	-0.5	-6.57	-5.19	-8.00	-6.40	-4.78	-3.43	-3.42	-2.84	-8.05	-6.69	-9.52	-8.09	-3.55	-2.96	-3.93	-3.37
	-0.8	-13.60	-12.36	-15.49	-14.50	-3.80	-3.05	-3.47	-2.90	-14.55	-13.60	-16.06	-15.27	-3.40	-2.68	-4.06	-3.48
500	0.0	-3.48	-2.95	-3.51	-3.06	-6.62	-2.89	-3.40	-2.84	-4.03	-3.47	-4.10	-3.54	-4.13	-3.48	-3.96	-3.40
	-0.5	-6.02	-4.70	-7.68	-5.85	-6.80	-3.44	-3.40	-2.83	-7.78	-6.27	-9.93	-8.03	-4.86	-3.33	-3.92	-3.39
	-0.8	-16.26	-14.23	-20.46	-18.73	-4.94	-4.25	-3.49	-2.93	-17.70	-16.28	-21.51	-20.43	-4.27	-3.36	-4.03	-3.47
1000	0.0	-3.41	-2.86	-3.44	-2.89	-10.13	-3.03	-3.39	-2.84	-3.99	-3.47	-4.08	-3.50	-4.24	-3.57	-3.93	-3.39
	-0.5	-5.66	-4.34	-7.17	-5.31	-6.98	-3.78	-3.41	-2.86	-7.23	-5.80	-9.34	-7.38	-8.69	-3.63	-3.96	-3.39
	-0.8	-18.76	-15.61	-25.69	-22.03	-10.26	-6.22	-3.43	-2.87	-21.47	-19.22	-28.55	-26.33	-6.22	-4.89	-4.05	-3.42

[1] Source: Schwert (1987a), Tables 1 and 2.

[2] MacKinnon's (1993) asymptotic critical values of the ADF t- test with a constant and no trend at a 1% (5%) significance level is -3.43 (-2.86). With a constant and a trend, it is -3.96 (-3.41).

Table A3.8 - Schwert's critical values of normalized bias unit root tests when the DGP is an IMA(1,1) with negative serial correlation: $\Delta y_t = \varepsilon_t + \theta\,\varepsilon_{t-1}$.[1,2]

Sample size T	MA parameter θ	Constant, no trend				Constant and trend			
		$Z(\rho_c)$ and l_4		$Z(\rho_c)$ and l_{12}		$Z(\rho_{ct})$ and l_4		$Z(\rho_{ct})$ and l_{12}	
		Size of the test		Size of the test		Size of the test		Size of the test	
		0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
25	0.0	-17.2	-12.4	-15.8	-11.0	-22.3	-17.7	-18.9	-14.0
	-0.5	-28.9	-24.4	-32.9	-26.3	-32.5	-28.8	-32.6	-27.3
	-0.8	-35.0	-31.5	-41.1	-35.1	-36.7	-33.6	-37.0	-31.9
50	0.0	-19.4	-13.8	-17.5	-12.8	-25.2	-20.3	-23.1	-17.4
	-0.5	-44.4	-35.1	-54.9	-43.2	-51.1	-44.2	-59.0	-49.5
	-0.8	-62.0	-55.9	-80.5	-70.1	-65.4	-59.6	-77.3	-66.7
100	0.0	-20.5	-14.2	-20.0	-14	-28.0	-21.6	-26.3	-20.3
	-0.5	-61.8	-44.3	-83.0	-62.8	-76.4	-62.4	-97.9	-80.4
	-0.8	-111.9	-100.7	-160.2	-142.2	-117.9	-108.8	-160.3	-142.9
250	0.0	-20.9	-14.4	-21.2	-14.8	-29.9	-22.4	-30.4	-22.9
	-0.5	-71.8	-46.6	-113.7	-74.6	-104.4	-75.1	-158.3	-118.2
	-0.8	-240.4	-211.1	-386.7	-345.6	-260.4	-239.7	-409.4	-373.1
500	0.0	-21.0	-14.4	-21.6	-14.7	-29.7	-22.2	-30.9	-23.2
	-0.5	-63.9	-39.4	-110.4	-64.5	-104.1	-69.1	-180.2	-120.3
	-0.8	-381.4	-306.9	-719.7	-617.3	-431.5	-382.8	-779.8	-713.3
1000	0.0	-20.6	-14.1	-21.2	-14.3	-29.8	-22.3	-31.2	-23.1
	-0.5	-59.2	-35.0	-97.1	-54.3	-94.6	-61.7	-164.4	-103.8
	-0.8	-576	-410.4	-1212.0	-900.5	-732.1	-604.6	-1460.0	-1262

[1] Source: Schwert (1987a), Tables 3 and 4.

[2] MacKinnon's asymptotic critical values of the normalized bias test with a constant and no trend at a 1% (5%) significance level is -20.6 (-14.1). With a constant and a trend, it is -29.4 (-21.7).

4. DYNAMIC ADJUSTMENT MODELS, LONG-RUN RELATIONSHIPS AND COINTEGRATION.

4.1 Introduction.

Even though it may be of interest to characterize separately the univariate behaviour of single time-series, ultimately macroeconomic theory attempts to draw the joint determination and causal links between time-series while an econometrician's task is to check whether these are compatible with the data. Since macroeconomics is concerned with explaining the time paths of variables over time, it is natural that dynamic specifications be of primary interest in empirical researches. Such specifications may arise either between only $I(0)$ variables or between $I(1)$ variables. In the latter case, a special long-run relationship is said to exist. Many new techniques can be used to investigate its existence and a good understanding of the econometric results specific to econometric models involving integrated series is required to characterize properly the link between many macroeconomic variables.

In this chapter, we will first present how to construct and estimate different parameterizations of dynamic models in the general case, which are both applicable to stationary and nonstationary time-series. Then, we will discuss issues related specifically to the econometric treatment of integrated regressors. Finally, we will generalize to systems of multivariate time-series processes.

4.2 Dynamic Models and Error-Correction Models.

A simple, yet very general dynamic specification is given by the autoregressive distributed lag, or $ADL(m, n; k)$, model between y_t and k regressors, $x_{1t}, x_{2t}, \ldots, x_{kt}$:

$$y_t = \kappa + \sum_{i=1}^{m} \phi_i \, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{n} \lambda_{ji} \, x_{jt-i} + \varepsilon_t \ , \tag{4.1}$$

where $\varepsilon_t \sim$ i.i.d.$(0, \sigma_\varepsilon^2)$, and $|\Sigma_{i=1}^{m} \phi_i| < 1$ is a stability condition on the model. Notice that n is the maximum lag length of the x_j's. For some j and some n_j, nothing prevents that $\lambda_{ji} = 0$ for $i \in [n_j, n\,]$. Using the lag operator, this equation can be represented as:

$$\Phi(L)\, y_t = \kappa + \sum_{j=1}^{k} \Lambda_j(L) x_{jt} + \varepsilon_t \ , . \tag{4.2}$$

where $\Phi(L) = 1 - \Sigma_{i=1}^{m} \phi_i L^i$ and $\Lambda_j(L) = \Sigma_{i=0}^{n} \lambda_{ji} L^i$, $\forall j = 1, \ldots, k$.

For such a model, provided joint stationarity of all its variables, and the absence of stochastic shocks in the long run, i.e. $E(\varepsilon_t) = 0$, then the long-run values of the variables are given by their unconditional expectations: $y^* = E(y_t)$, $x_j^* = E(x_{jt})$, for all $j = 1, \ldots, k$ and for all t.[1] Consequently, the long-run equilibrium between y_t, x_{1t}, x_{2t}, ..., and x_{kt} is such that:

$$\Phi(L)y^* = \kappa + \sum_{j=1}^{k} \Lambda_j(L)x_j^* \quad \Leftrightarrow \quad \Phi(1)y^* = \kappa + \sum_{j=1}^{k} \Lambda_j(1)x_j^*$$

where $\Phi(1) = 1 - \Sigma_{i=1}^{m} \phi_i$. This, in turn, implies that:

$$y^* = \frac{\kappa}{\Phi(1)} + \sum_{j=1}^{k} \frac{\Lambda_j(1)}{\Phi(1)} x_j^* = \left(\frac{\kappa}{1 - \Sigma_{i=1}^{m} \phi_i}\right) + \sum_{j=1}^{k} \left(\frac{\Sigma_{i=0}^{n} \Lambda_{ji}}{1 - \Sigma_{i=1}^{m} \phi_i}\right) x_j^* \ . \tag{4.3a}$$

or

$$y^* = \beta_0 + \sum_{j=1}^{k} \beta_j x_j^* \tag{4.3b}$$

A parameter $\beta_j \equiv \partial y^*/\partial x_j^*$ is called the long-run multiplier of y_t with respect to x_{jt}.

This general dynamic model can be subjected to many equivalent linear transformations that do not affect the properties of the models but that help emphasizing different sets of coefficients depending upon the interest of one's quest. While these transformations only involve very basic algebra, I find that one's intuition is considerably enhanced by considering them in turn, in increasing order of complexity from the simpler to the more general case. Tables 4.1 and 4.2 guide us through many of these transformations when considering representations of an *ADL(1,1;1)*, and an *ADL(2,2,;1)* that are respectively given by:

$$y_t = \kappa + \phi_1 y_{t-1} + \lambda_0 x_t + \lambda_1 x_{t-1} + \varepsilon_t \ , \tag{4.4}$$

$$y_t = \kappa + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \lambda_0 x_t + \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \varepsilon_t \ . \tag{4.5}$$

We now discuss the intuition, the relative interest, as well as the estimation and inference issues related to some of these transformations.

Transformations #2 (Bårdsen's) and #3 in Tables 4.1 and 4.2 allow us to estimate directly the sum of the λ_i's and to retrieve the standard-error of this sum from an OLS regression output.

---

[1] As it will be discussed later, the hypothesis of joint stationarity of the variables is not always required for the existence of a long-run stationary equilibrium between several variables. Some of these variables may be integrated and thus nonstationary while there exists cointegrating relationships between them.

Transformations #4 (a, b, and c) and #5 (a and b) provide a different reparameterization that has become increasingly popular in empirical works in the recent years. These are called error-correction models (ECM) and they offer a useful strategy for modelling statistically the short-run dynamics without neglecting a consistent link with the long-run relationship between these variables. For instance, transformation # 5a, that is rewritten here for convenience:

$$\begin{aligned}\Delta y_t &= \kappa + [\phi_1 + \phi_2 - 1](y_{t-1} - \beta_1 x_{t-1}) - \phi_2 \Delta y_{t-1} + \lambda_0 \Delta x_t - \lambda_2 \Delta x_{t-1} + \varepsilon_t \\ &= \kappa + \alpha (y_{t-1} - \beta_1 x_{t-1}) + \varphi_{11} \Delta y_{t-1} + \varphi_{02} \Delta x_t + \varphi_{12} \Delta x_{t-1} + \varepsilon_t , \qquad (4.6)\end{aligned}$$

expresses changes in y_t as the sum of the contributions from the correction needed to compensate for a lagged deviation from the long-run equilibrium, and from the dynamic adjustments, captured by lagged changes in y_t and contemporaneous and lagged changes in x_t. Indeed, whenever the mean-corrected value of y_{t-1} is bigger than the mean-corrected value of $\beta_1 x_{t-1}$, α ought to be negative to correct the short-run deviation from the long-run equilibrium, by bringing a lower value of Δy_t.[2] The coefficient α in the error-correction term $\{\alpha [y_{t-1} - \beta_1 x_{t-1}]\}$ is therefore the proportion of the short-run deviation in *t-1* that is offset by a movement in y_t. Joint estimation of the short-run dynamic parameters and of the long-run relationship coefficient between y_t and x_t can be carried out with a simple non-linear least-squares procedure.

Transformations #4a, #4b and #4c (or #5a, and #5b) illustrate that there is more than one way to specify an ECM representation. First, transformations #4a, #4b and #4c show that it is possible to enter the ECM term as a function of the difference between lagged values of y_t and x_t. In these instances, however, one must be sure to include the relevant lagged values of the variables in level so that the representation does not falsely impose *a priori* that β_1 be equal to one without testing. Secondly, transformations #5a and #5b illustrate that it is always possible to write a correct ECM with a lag on the error-correction term of whatever order. In practice, this value is never chosen to be greater than the maximum lag on either Δy_t or Δx_t plus one. Considering an *ADL(m,n;1)*, such as:

---

[2] The presence of an intercept in equation (4.6) is equivalent to the use of all the mean-corrected variables. On the basis of the long-run relationship between y_t and x_t given by $y_t = \beta_0 + \beta_1 x_t$, one representation equivalent to equation (4.6) would be $\Delta y_t = \alpha [y_{t-1} - \beta_0 - \beta_1 y_{t-1}] + \varphi_{11} \Delta y_{t-1} + \varphi_{02} \Delta x_t + \varphi_{12} \Delta x_{t-1} + \varepsilon_t$, since $\kappa = -\beta_0 \alpha$. It is also possible that κ plays the role of a constant inherent to the dynamics.

$$y_t = \kappa + \sum_{i=1}^{m} \phi_i y_{t-i} + \sum_{i=0}^{n} \lambda_i x_{t-i} + \varepsilon_t \,, \tag{4.7}$$

in order to estimate a corresponding ECM, one could assume to begin with that *m* and *n* are both equal to *p* so that

$$\begin{aligned}\Delta y_t &= \kappa + \sum_{i=1}^{p-1} \left[\sum_{l=1}^{i} \phi_l - 1\right] \Delta y_{t-i} + \sum_{i=0}^{p-1} \left[\sum_{l=0}^{i} \lambda_l\right] \Delta x_{t-i} + \left[\sum_{l=1}^{p} \phi_l - 1\right] (y_{t-p} - \beta_1 x_{t-p}) + \varepsilon_t \,, \\ &= \kappa + \sum_{i=1}^{p-1} \varphi_{i1} \Delta y_{t-i} + \sum_{i=0}^{p-1} \varphi_{i2} \Delta x_{t-i} + \alpha \, (y_{t-p} - \beta_1 x_{t-p}) + \varepsilon_t \,. \end{aligned} \tag{4.8}$$

If this ECM is an appropriate representation of the data that is consistent with an existing long-run relationship between y_t and x_t, the coefficient α of the error-correction term should be negative and significant. As discussed in Wickens and Breusch (1988), too small a value of *p* can be reflected by left-in serial correlation in the residuals. On the other hand too large a value of *p* may arise in the case of a misspecified error-correction term with a non-significant coefficient but with significant estimates of the $\varphi_{p+1,1}$ and $\varphi_{p+1,2}$, etc. It is important to notice that the appropriate specification of equation (4.8) cannot be obtained by eliminating non significant lags. Because these coefficients are themselves functions of the partial cumulative sum of the *ADL* coefficients, one must test whether the coefficients of Δy_{t-p+1} and Δx_{t-p+1} are jointly different from those on Δy_{t-p+2} and Δx_{t-p+2}, etc. to check the relevance of a higher maximum lag order. In the same vein, the individual examination of the coefficients on either Δy_{t-p+1} and Δx_{t-p+1} provide a way of testing whether $m=n$ as it has been imposed at the outset in the general model.

There are other instances in which one may be interested mainly in estimating the long-run multiplier. The Bewley transformation of the *ADL* representation, illustrated as transformations #6 in Table 1, or #6a and #6b in Table 2, respectively for an *ADL*(1,1;2) and an *ADL*(2,2;2), suggests a regression equation that permits to estimate directly the long-run multiplier β_1 relating y_t to x_t.[3] By regressing the dependent variable on a constant, x_t, and a sequence of differences of increasing order in both y_t and x_t, it is possible to estimate the long-run multiplier while statistically preserving the short-run dynamics in the empirical equation. It is also equivalent to specify the Bewley specification as a regression of y_t on a constant, x_t, and a sequence of lagged-first differences in both y_t and x_t. Referring to the *ADL(m,n;*1) above, this latter Bewley representation would amount to:

---

[3] Notice that while it is often the case that one can provide a particular economic interpretation for the long-run multipliers, the coefficients of the regressors capturing the dynamics of adjustment do not generally have a specific meaning.

$$y_t = \beta_0 + \beta_1 x_t + \sum_{i=0}^{m-1} \eta_i \Delta y_{t-i} + \sum_{j=1}^{n-1} \psi_j \Delta x_{t-j} + u_t \, . \tag{4.9}$$

The different equivalent representations of a dynamic relationship and the issues regarding the appropriate econometric treatment of these empirical equations extend directly to the *ADL(m,n;k)* case, involving *k* regressors. Table 4.3 summarizes these transformations.

Because Δy_t (i.e. y_t-y_{t-1}) appears amongst the regressors of a Bewley representation, the correlation between this regressor and the error term could make the OLS estimator of the β_j's inconsistent. In this case, the use of the generalized instrumental variable estimator would give consistent estimator of the long-run multipliers. The formulation of equation (4.9) then suggests the use of all the regressors in the *ADL* representation (i.e. the *m* lags on y_t and the *n* lags on x_t) *as* instruments for the regressors in equation (4.9). As shown by Wickens and Breusch (1988), in this case, the GIVE estimator will directly give the exact numerical estimate of the long-run multiplier that would have been reckoned from the OLS estimates of the coefficients in the *ADL* representation. Another advantage of the Bewley equation is that its estimation of the long-run multiplier may be more robust to an overparameterization of the lag structure and to the neglect of restriction on the lag structure than it would have been the case from an estimation of its equivalent ECM representation. In any case, statistical adequacy checks remain in order.

The different representations of a dynamic relationship amount to fairly ordinary empirical equations requiring the same econometric care that has been discussed throughout this guide. One issue has not been addressed explicitly and deserves further discussion. Namely, for consistent and efficient estimation of the parameters of a well-specified individual equation, and for valid inference, the regressors of this conditional expectation equation, or the instruments when applicable, must be weakly exogenous with respect to the parameters of interest.[4] That is, as defined by Engle, Hendry and Richard (1983), conditional inference on individual equation parameters need not involve any loss of information that could have been prevented from estimating jointly a system of equations for y_t and the other regressors.[5] We will come back later to the issues of exogeneity, in particular when dealing more specifically with I(1) variables and with system of equations. In the context of dynamic models with I(0) variables, it is important to keep in mind that both short-run and long-run coefficients in an individual dynamic equation with endogenous regressors will be incorrectly estimated.

---

[4] While weak exogeneity matters for efficient estimation and inference, there are other concepts of exogeneity such as strong- and super- exogeneity as exposed in Engle, Hendry and Richard (1983). A variable x_t is said to be strongly exogenous if it is weakly exogenous and if it is not Granger caused by y_t. The latter concept only means that the determination of the current value of x_t is not related to any lagged value of y_t. Strong exogeneity is required for multi-step forecasts of a variable y_t in the future based on conditional forecasts of future x_t. Superexogeneity relates to conditions that would prevent the Lucas critique to apply for economic policy evaluation based on econometric model.

The dynamic models we have just discussed are relevant both for the case of all stationary regressors but also in special cases with integrated regressors. In the latter case, some special issues deserve further discussions, which we carry in the next sections. It is however fundamental to remember that, as any empirical equations, dynamic models must satisfy the set of conventions discussed in chapter 1. The adequacy of the specification, the behaviour of the residuals, the stability of the model, and so forth, all underlie the validity of the estimation procedure. Possible departures from these conventions can yield seriously spurious inference and, if ignored, they can therefore invalidate the drawn conclusions. One should therefore apply a battery of adequacy tests to validate the statistical representation of the data before entertaining tests involving some economic interpretations.

---

[5] Weak exogeneity of x_t with respect to its coefficient in a dynamic conditional expectation equation for y_t requires that the error term of this dynamic equation has a zero long-run covariance with the innovation of x_t. If the lagged innovations of y_t Granger cause the innovation of x_t, then x_t is not weakly exogenous. Weak exogeneity of a regressor with respect to its parameter also fails to hold when theory suggests that the regressor should be an unobserved expectational variable, x^*_t, but that it is replaced by the realized value of this variable, say x_t. In this case, even if the error term of the y_t equation was contemporaneously uncorrelated with the error term of the forecast equation for x_t, x_t would not be weakly exogenous with respect to its coefficient in the y_t equation. (See Banerjee *et al.* 1993). In both instances, the alternative for $I(0)$ variables would be to estimate jointly the system of dynamic equations.

Table 4.1 - Linear transformations of an autoregressive distributed lag ADL(*1,1;1*) model.[1, 2]

	Model	Long-run equilibrium
	$y_t = \kappa + \phi_1 y_{t-1} + \lambda_0 x_t + \lambda_1 x_{t-1} + \varepsilon_t$ (4.4)	$\beta_0 = \kappa / (1-\phi_1)$ $\beta_1 = (\lambda_0 + \lambda_1) / (1-\phi_1)$
Transformation # 1	Subtract y_{t-1} from both sides of (4.4). $\Delta y_t = \kappa + [\phi_1 - 1] y_{t-1} + \lambda_0 x_t + \lambda_1 x_{t-1} + \varepsilon_t$	
Transformation # 2	The Bårdsen transformation: Transformation #1 + add and subtract $\lambda_0 x_{t-1}$ on the RHS. $\Delta y_t = \kappa + [\phi_1 - 1] y_{t-1} + \lambda_0 \Delta x_t + [\lambda_0 + \lambda_1] x_{t-1} + \varepsilon_t$	
Transformation # 3	Transformation #1 + add and subtract $\lambda_1 x_t$ on the RHS. $\Delta y_t = \kappa + [\phi_1 - 1] y_{t-1} - \lambda_1 \Delta x_t + [\lambda_0 + \lambda_1] x_t + \varepsilon_t$	
Transformation # 4	Error-correction model (ECM) I : Transformation #2 + add and subtract $[(\lambda_1 - 1) x_{t-1}]$ on the RHS. $\Delta y_t = \kappa + [\phi_1 - 1] (y_{t-1} - x_{t-1}) + \lambda_0 \Delta x_t + [\lambda_0 + \lambda_1 + \phi_1 - 1] x_{t-1} + \varepsilon_t$	
Transformation # 5	Error-correction model (ECM) II: Transformation #4 + add and subtract $[(\phi_1 - 1) \beta_1 x_{t-1}] \equiv [-(\lambda_0 + \lambda_1) x_{t-1}]$ on the RHS. $\Delta y_t = \kappa + [\phi_1 - 1] (y_{t-1} - \beta_1 x_{t-1}) + \lambda_0 \Delta x_t + \varepsilon_t$	
Transformation # 6	The Bewley transformation: Subtract $[\phi_1 y_t]$ from both sides of (4.4) + add and subtract $[\lambda_1 x_t]$ on the RHS. $y_t = \varsigma\kappa + \beta_1 x_t - \varsigma\phi_1 \Delta y_t - \varsigma\lambda_1 \Delta x_t + \varsigma\, \varepsilon_t$	

[1] NOTE: $\Delta y_{t} = y_{t} - y_{t-1}$; $\Delta x_{t} = x_{t} - x_{t-1}$; and $\zeta = [1 - \phi_1]^{-1}$

[2] The long-run equilibrium is such that $y_s = y^*$ and $x_s = x^*$ for all s.

Table 4.2 - Linear transformations of an autoregressive distributed lag ADL(*2,2;1*) model.[1,2]

<table>
<tr><th colspan="2">Model</th><th>Long-run equilibrium</th></tr>
<tr><td colspan="2">$y_t = \kappa + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \lambda_0 x_t + \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \varepsilon_t$ (4.5)</td><td>$\beta_0 = \frac{\kappa}{(1-\phi_1-\phi_2)}$
$\beta_1 = \frac{\lambda_0+\lambda_1+\lambda_2}{(1-\phi_1-\phi_2)}$</td></tr>
<tr><td rowspan="2">Transformation # 1</td><td colspan="2">Subtract y_{t-1} from both sides of (4.5) + add and subtract $[(\phi_1 - 1)\, y_{t-1}]$ on the RHS.</td></tr>
<tr><td colspan="2">$\Delta y_t = \kappa + [\phi_1 - 1]\,\Delta y_{t-1} + (\phi_1 + \phi_2 - 1)\,y_{t-2} + \lambda_0 x_t + \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \varepsilon_t$</td></tr>
<tr><td rowspan="2">Transformation # 2</td><td colspan="2">The Bårdsen transformation:
Transformation #1 + add and subtract $[\lambda_0 x_{t-1} + \lambda_0 x_{t-2} + \lambda_1 x_{t-2}]$ on the RHS.</td></tr>
<tr><td colspan="2">$\Delta y_t = \kappa + [\phi_1 - 1]\,\Delta y_{t-1} + [\phi_1 + \phi_2 - 1]\,y_{t-2}$
$+ \lambda_0 \Delta x_t + [\lambda_0 + \lambda_1]\Delta x_{t-1} + [\lambda_0 + \lambda_1 + \lambda_2]\,x_{t-2} + \varepsilon$</td></tr>
<tr><td rowspan="2">Transformation # 3</td><td colspan="2">Transformation #1 + add and subtract $[\lambda_1 x_t + \lambda_2 x_t]$ on the RHS.</td></tr>
<tr><td colspan="2">$\Delta y_t = \kappa + [\phi_1 - 1]\,\Delta y_{t-1} + [\phi_1 + \phi_2 - 1]\,y_{t-2}$
$+ [\lambda_0 + \lambda_1 + \lambda_2]\,x_t - \lambda_1 \Delta x_t - \lambda_2 \Delta_2 x_t + \varepsilon_t$</td></tr>
<tr><td rowspan="2">Transformation # 4a</td><td colspan="2">Error-correction model (ECM) Ia :
Transformation #2 + add and subtract $[(\phi_1 - 1)\, x_{t-1} + \phi_2\, x_{t-2}]$ on the RHS.</td></tr>
<tr><td colspan="2">$\Delta y_t = \kappa + [\phi_1 - 1]\,(y_{t-1} - x_{t-1}) + \phi_2\,(y_{t-2} - x_{t-2}) + \lambda_0 \Delta x_t$
$+ [\lambda_0 + \lambda_1 + \phi_1 - 1]\,x_{t-1} + [\phi_2 + \lambda_2]\,x_{t-2} + \varepsilon_t$</td></tr>
<tr><td rowspan="2">Transformation # 4b</td><td colspan="2">Error-correction model (ECM) Ib :
ECM Ia + add and subtract $[(\phi_2 + \lambda_2)\, x_{t-1} + \phi_2\, y_{t-1}]$ on the RHS.</td></tr>
<tr><td colspan="2">$\Delta y_t = \kappa + [\phi_1 + \phi_2 - 1]\,(y_{t-1} - x_{t-1}) - \phi_2\,\Delta y_{t-1} + \lambda_0 \Delta x_t$
$+ [\lambda_0 + \lambda_1 + \lambda_2 + \phi_1 + \phi_2 - 1]\,x_{t-1} - \lambda_2\,\Delta x_{t-1} + \varepsilon_t$</td></tr>
<tr><td rowspan="2">Transformation # 4c</td><td colspan="2">Error-correction model (ECM) Ic :
ECM Ia + add and subtract $[(\phi_1 - 1)\,(y_{t-2} - x_{t-2}) + (\lambda_0 + \lambda_1)\, x_{t-2}]$ on the RHS.</td></tr>
<tr><td colspan="2">$\Delta y_t = \kappa + [\phi_1 - 1]\,\Delta y_{t-1} + [\phi_1 + \phi_2 - 1]\,(y_{t-2} - x_{t-2}) + \lambda_0 \Delta x_t$
$+ [\lambda_0 + \lambda_1]\,\Delta x_{t-1} + [\lambda_0 + \lambda_1 + \lambda_2 + \phi_1 + \phi_2 - 1]\,x_{t-2} + \varepsilon_t$</td></tr>
</table>

Table 4.2 - Linear transformations of an autoregressive distributed lag ADL(*2,2;1*) model.[1, 2]

Transformation # 5a	Error-correction model (ECM) II: Transformation #4b + add and subtract $[(\phi_1 + \phi_2 - 1)\, \beta_1\, x_{t-1}] \equiv [-(\lambda_0 + \lambda_1 + \lambda_2)\, x_{t-1}]$ on the RHS. $\Delta y_t = \kappa + [\phi_1 + \phi_2 - 1](y_{t-1} - \beta_1 x_{t-1}) - \phi_2 \Delta y_{t-1} + \lambda_0 \Delta x_t - \lambda_2 \Delta x_{t-1} + \varepsilon_t$
Transformation # 5b	Error-correction model (ECM) II: Transformation #4c + add and subtract $[(\phi_1 + \phi_2 - 1)\, \beta_1\, x_{t-2}] \equiv [-(\lambda_0 + \lambda_1 + \lambda_2)\, x_{t-2}]$ on the RHS. $\Delta y_t = \kappa + [\phi_1 - 1]\, \Delta y_{t-1} + [\phi_1 + \phi_2 - 1]\, (y_{t-2} - \beta_1 x_{t-2}) + \lambda_0 \Delta x_t + [\lambda_0 + \lambda_1] \Delta x_{t-1} + \varepsilon_t$
Transformation # 6a	The Bewley transformation I: Subtract $[(\phi_1 + \phi_2)\, y_t]$ from both sides of (4.5) + add and subtract $[(\lambda_1 + \lambda_2)\; x_t]$ on the RHS. $y_t = \zeta\kappa + \beta_1 x_t - \zeta\phi_1 \Delta y_t - \zeta\phi_2 \Delta_2 y_t - \zeta\lambda_1 \Delta x_t - \zeta\lambda_2 \Delta_2 x_t + \zeta\varepsilon_t$
Transformation # 6b	The Bewley transformation II: Subtract $[(\phi_1 + \phi_2)\, y_t]$ from both sides of (4.5) + add and subtract $[(\lambda_1 + \lambda_2)\; x_t] + [\lambda_2 x_{t-1} + \phi_2 y_{t-1}]$ on the RHS . $y_t = \zeta\kappa + \beta_1 x_t - \zeta[\phi_1 + \phi_2] \Delta y_t - \zeta\, \phi_2 \Delta y_{t-1} - \zeta\, [\lambda_1 + \lambda_2] \Delta x_t - \zeta \lambda_2 \Delta x_{t-1} + \zeta\varepsilon_t$

[1] NOTE: $\Delta y_{t-i} = y_{t-i} - y_{t-i-1}$; $\Delta x_{t-i} = x_{t-i} - x_{t-i-1}$; $\Delta_s y_{t-i} = y_{t-i} - y_{t-i-s}$; $\Delta_s x_{t-i} = x_{t-i} - x_{t-i-s}$; and $\zeta = [1 - \phi_1 - \phi_2]^{-1}$.

[2] The long-run equilibrium is such that $y_s = y^*$ and $x_s = x^*$ for all s.

Table 4.3 - Linear transformations of an autoregressive distributed lag ADL(*m,n;k*) model.[1, 2]

	Model	Long-run equilibrium
	$y_t = \kappa + \sum_{i=1}^{m} \phi_i\, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{n} \lambda_{ji}\, x_{jt-i} + \varepsilon_t$	$\beta_0 = \left(\frac{\kappa}{1-\sum_{i=1}^{m} \phi_i}\right)$ $\beta_j = \left(\frac{\sum_{i=0}^{n} \lambda_{ji}}{1-\sum_{i=1}^{m} \phi_i}\right)$
Bårdsen Representation	$\Delta y_t = \kappa + \sum_{i=1}^{m-1} \left[\sum_{l=1}^{i} \phi_l - 1\right] \Delta y_{t-i} + \sum_{j=1}^{k} \left(\sum_{i=0}^{n-1} \left[\sum_{l=0}^{i} \lambda_{jl}\right] \Delta x_{jt-i}\right) + \left[\sum_{l=1}^{m} \phi_l - 1\right] y_{t-m} + \sum_{j=1}^{k} \left[\sum_{l=0}^{n} \lambda_{jl}\right] x_{jt-n} + \varepsilon_t$	
ECM Representation I (setting *m=n=p*)	$\Delta y_t = \kappa + \sum_{i=1}^{p-1} \left[\sum_{l=1}^{i} \phi_l - 1\right] \Delta y_{t-i} + \sum_{j=1}^{k} \left(\sum_{i=0}^{p-1} \left[\sum_{l=0}^{i} \lambda_{jl}\right] \Delta x_{jt-i}\right)$ $+ \left[\sum_{l=1}^{p} \phi_l - 1\right] \left(y_{t-p} - \sum_{j=1}^{k} x_{jt-p}\right) + \sum_{j=1}^{k} \left[\sum_{l=0}^{p} \lambda_{jl} + \sum_{l=1}^{p} \phi_l - 1\right] x_{jt-p} + \varepsilon_t$	
ECM Representation II (setting *m=n=p*)	$\Delta y_t = \kappa + \sum_{i=1}^{p-1} \left[\sum_{l=1}^{i} \phi_l - 1\right] \Delta y_{t-i} + \sum_{j=1}^{k} \left(\sum_{i=0}^{p-1} \left[\sum_{l=0}^{i} \lambda_{jl}\right] \Delta x_{jt-i}\right)$ $+ \left[\sum_{l=1}^{p} \phi_l - 1\right] \left(y_{t-p} - \sum_{j=1}^{k} \beta_j x_{jt-p}\right) + \varepsilon_t$	

Table 4.3 - Linear transformations of an autoregressive distributed lag ADL(*m,n;k*) model.[1, 2]

Bewley Representation I	$y_t = \zeta\kappa + \sum_{j=1}^{k}\beta_j x_{jt} - \left[\zeta\sum_{l=1}^{m}\phi_l\right]\Delta_l y_t - \sum_{j=1}^{k}\left[\zeta\sum_{l=0}^{n}\lambda_{jl}\right]\Delta_l x_t + \zeta\varepsilon_t$
Bewley Representation II	$y_t = \zeta\kappa + \sum_{j=1}^{k}\beta_j x_{jt} - \sum_{i=0}^{m-1}\left[\zeta\sum_{l=i+1}^{m}\phi_l\right]\Delta y_{t-i} - \sum_{j=1}^{k}\left(\sum_{i=0}^{n-1}\left[\zeta\sum_{l=i+1}^{n}\lambda_{jl}\right]\Delta x_{jt-i}\right) + \zeta\varepsilon_t$

[1] NOTE: $\Delta y_{t-l} = y_{t-l} - y_{t-l-1}$; $\Delta x_{t-l} = x_{t-l} - x_{t-l-1}$; $\Delta_s y_{t-l} = y_{t-l} - y_{t-l-s}$; $\Delta_s x_{t-l} = x_{t-l} - x_{t-l-s}$; and $\zeta = [1 - \Sigma^{m}_{l=1}\phi_l]^{-1}$.

[2] The long-run equilibrium is such that $y_s = y^*$ and $x_{js} = x^*$ for all s.

4.3 Relationships Between Integrated Series In Individual Equations: Spurious Regression vs. Cointegration.

4.3.1 Spurious regressions in econometrics

It is quite frequent to find applied papers reporting empirical results for a regression equation such as:

$$y_t = \mathbf{x}_t'\,\beta + u_t, \qquad (4.10)$$

where $\mathbf{x}_t' = [1\ x_{2t}\,x_{3t} \dots x_{kt}]$ is of dimension $k \times 1$, that is characterized by a reportedly high degree of fitness as reflected by large values of R^2 and $\bar{R}^2$, but with a very low Durbin-Watson (DW) statistics.

For instance, let us consider the cases of two time-series y_t and x_t that are both integrated of order one with drift. As the sample size grows, the trend in the series will dominate in their movements, thus leading to an R^2 that approaches unity.[6] Moreover, if this relationship is spurious, the error term u_t is nothing but the linear combination of two integrated variables that in general would be nonstationary as well, i.e. $u_t \sim I(1)$ or $u_t = \rho\, u_{t-1} + \varepsilon_t$ with $\rho = 1$. Since the Durbin-Watson statistic is approximately given by $DW \approx 2 - 2\hat{\rho}$, then $DW \to 0$. In fact, when a DW-statistic is low with high R^2, this often signals problems with the relationship being investigated. The same intuition would carry in the case of k regressors for which a subset is nonstationary when the relationship is spurious.[7]

---

[6] Technically, Phillips (1986) showed that the R^2 converges to a random variable in this case.

[7] Sargan and Bhagava (1983) have tabulated critical values of the DW statistic as a non-cointegration test for a few specific DGP's of the error terms. However, the sensitivity of the critical values to the true DGP limits its usefulness.

Suppose that we wish to test the null hypothesis H_0: $\beta_2 = \beta_3 = ... = \beta_k = 0$ in regression (4.10), the F-statistic version of a Wald test could be computed as:

$$F = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} \sim F(k-1, T-k) \text{ under } H_0, \tag{4.11}$$

according to usual econometric practices. If the null hypothesis is true in the population, then $u_t = y_t - x_t' \beta \sim I(1)$, since $y_t \sim I(1)$. This procedure would be at fault if the error term u_t is integrated because the correct distribution of the F-statistic is not the conventional F-distribution in this case. Indeed, the F-distribution under H_0 is valid only for a white noise, and therefore stationary error-term. Furthermore, with the F-test, the existence of stochastic trends in y_t and in at least some of the x_{it} would inflate the R^2 and raise the value of the F-statistic. This will make easier the rejection of the null hypothesis even if it is true. The F-test and other conventional tests are therefore biased towards finding spurious evidence of relationships when considering integrated regressors.

Monte Carlo experiments conducted by Granger and Newbold (1986) gave results compatible with those obtained by Nelson and Kang (1984). They also showed that with time series whose DGP is described either as random walks or as $IMA(1,1)$, the probability of finding a spurious relationship is very high (greater than 64% with tests conducted at the 5% significance level). Also, this probability increases with k as y_t is regressed on a constant and k-1 independent integrated variables.

Because many macroeconomic variables appear to be integrated or near-integrated, the possibility of finding spurious relationship should not be taken too lightly. However, there are some instances for which there may be valid relationships between integrated variables. We now turn to issues relevant for integrated variables.

4.3.2 Cointegrating relationships

Generally, a linear combination of two or more integrated series, say of order one, would yield an integrated series of the same order. Interestingly, however, some linear combinations may be integrated of a lesser order. Hence, while some series might be wandering apart in the short run, there may be some long run equilibrium relationships such that these variables move together over a longer horizon. Notice that short run deviations from a long-run equilibrium do not mean that the short-run dynamic of a particular economic phenomenon corresponds to the behaviour of economic agents being in disequilibrium. It only means that the short-run economic equilibrium may differ from the economic equilibrium outcome in the long run.

Formally, for $\mathbf{x}_t$ being a vector of dimension $N \times 1$ whose elements are $I(d)$, with $d > 0$, the series in $\mathbf{x}_t$ are said to be cointegrated of order (d,b), denoted as $CI(d,b)$, if there exists at least one cointegrating vector, β_i of dimension $N \times 1$, such that $\beta_i' \mathbf{x}_t \sim I(d-b)$, for $d \geq b > 0$ and $\beta_i \neq 0_{N \times 1}$. In applied work, the most encountered case concerns $I(1)$ variables for which there exists a $CI(1,1)$ relationship, i.e. $\beta_i' \mathbf{x}_t \sim I(0)$. Notice that a cointegrating vector is identifiable only up to a scale of transformation since for any constant $c \neq 0$, $c\,\beta_i' \mathbf{x}_t \sim I(0)$ as well. Furthermore, for a vector $\mathbf{x}_t$ made up of N variables, there may exist up to $N\text{-}1$ linearly independent cointegrating vectors. If there are $r \leq N\text{-}1$ cointegrating vectors, the $N \times r$ cointegrating matrix $\beta = [\beta_1, \ldots, \beta_r]$ is such that the r elements of the vector $\beta' \mathbf{x}_t$ are trend-stationary. Another terminology related to cointegration is that if there are r cointegrating vectors that remove the random walks of a system of N integrated series, there are in fact $n\text{-}r$ common stochastic trends between these variables.[8] If there is no stationary long run relationship between the N integrated variables, then any estimated statistical equation between these variables is spurious.

Furthermore, a direct implication of cointegrating versus spurious relationships between $I(1)$ time series is that a regression equation can make sense only if it is "balanced", i.e. in the case of $I(1)$ variables whether the residuals of the empirical equation are stationary. This condition can apply in any of the following cases involving:

1. a dependent variable and regressors that are all $I(0)$;
2. an $I(1)$ dependent variable that is cointegrated with the $I(1)$ regressors, despite the presence of other $I(0)$ regressors in the equation;
3. an $I(0)$ dependent variable, with $I(1)$ regressors that are cointegrated together, despite the presence of other $I(0)$ regressors in the equation.[9]

---

[8] King, Plosser, Stock and Watson (1991), and Watson (1993) have shown how one can write a common stochastic trend representation of the system relating the N variables that form $\mathbf{x}_t$.

[9] Another example in the same vein could consist of a log-linear long-run nominal money demand equation specified in terms of the following variables: the money stock, m_t, real output, y_t, the price level, p_t, and the nominal interest rate, R_t

Econometric practice in the 1970's suggested by Granger and Newbold (1974) in the Box and Jenkins (1970) tradition would have lead automatically to differencing integrated variables before searching for a relationship between their first differences. The recent developments in macroeconometrics clearly indicate that this would be a mistake since we would then omit from the start the possibility of long-run relationships across individually nonstationary variables. Economic theory sometimes suggests long-run relationships (e.g. real interest rate parity, purchasing power parity, the quantity theory of money, balanced growth models, interest rate term structure relationships). While evidence of cointegration will not necessarily imply support for a single theory, it is nonetheless often a necessary condition for consistency between the data and some economic interpretations as well as part of the characteristics of the data to be explained.[10] We need therefore to discuss how we can estimate and test for the existence of cointegration.

4.3.3 Regressions and cointegration in individual equations

4.3.3.1 The case of two variables.

4.3.3.1.1 Estimation in static regressions

Let us consider two series x_{1t} and x_{2t} that are both integrated of order one, i.e. $I(1)$.[11] In this case, since $N = 2$, if there exists an independent cointegrating vector β, it is unique up to a scale of transformation. Therefore,

$$\beta' \mathbf{x}_t = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \beta_1 x_{1t} + \beta_2 x_{2t} = \upsilon_{1t} \sim I(0). \qquad (4.12)$$

If we normalize the cointegrating relationship above with respect to β_1 and we isolate x_{1t}, we obtain:

Suppose that both m_t and p_t are $I(2)$ but that they are $CI(2,1)$ in a way that real money demand is $I(1)$, and that y_t and R_t are both $I(1)$. If $[m_t\text{-}p_t]$, y_t and R_t are $CI(1,1)$, the equation with m_t as regressand and p_t as regressor would be balanced, as would an equation with m_t as regressand and all of p_t, R_t and y_t as regressors.

[10] For instance, suppose that short-term and long-term nominal interest rates are each integrated of order one. The expectation theory of the term structure says that the long-term interest rate is an average of the current and future short-term interest rates. Hence, evidence of cointegration between short- and long-term interest rate would be compatible with this theory. However, cointegration only means that any deviation from the long-run relationship is stationary. Therefore, a term structure theory with stationary risk premium or one alternative theory in which economic agents overdiscount the future would also be consistent with evidence of cointegration. (See Miron (1991)).

[11] As it is implied from the definition of cointegration itself, a first necessary step is to establish that at least some of the variables of interest are individually integrated to begin with. That is why, at the early stage of an empirical research, it is advisable to characterize the data at hands with the tools presented in Chapter 3.

$$x_{1t} = \xi x_{2t} + u_{1t} \tag{4.13}$$

where $\xi \equiv -\beta_2/\beta_1$, and $u_{1t} = -\upsilon_t/\beta_1$. Equation (4.13) therefore amounts to a static regression model.[12] Hence, x_{1t} and x_{2t} are $CI(1,1)$ if and only if u_{1t} is $I(0)$, while they are not cointegrated if u_{1t} is $I(1)$. In the latter case, equation (4.13) gives rise to a spurious relationship as discussed in the previous section.

Furthermore, this empirical representation can be estimated by OLS, as Stock (1987) has shown that the OLS estimator of the coefficient of an integrated regressor such as ξ on x_{2t} is theoretically superconsistent. That is, the OLS estimator $\hat{\xi}$ tends even more quickly towards its true value as the sample size T grows than it would have been the case if its associated regressor had been $I(0)$.[13] This is particularly important since equation (4.13) does not model explicitly any dynamics that may exist between x_{1t} and x_{2t} that is left in the error term. Indeed, cointegration only means that u_{1t} is $I(0)$ not that it is serially uncorrelated. Yet, superconsistency implies that the OLS estimator of ξ, denoted as $\hat{\xi}$, will tend very fast towards its true value with a growing sample even if there is missing dynamic or simultaneity bias.

This can be intuitively understood when considering that the bias vanishes for a large enough sample, i.e.

$$\lim_{T\to\infty} E(\hat{\xi}-\xi \,|\, x_{2t}) = \lim_{T\to\infty} E(\,(x_{2t}'x_{2t})^{-1}x_{2t}'u_{1t} \,|\, x_{2t}) = 0 \;. \tag{4.14}$$

Since $x_{2t} \sim I(1)$, its variance explodes with T, so that $(x_{2t}'x_{2t})^{-1}$ tends much faster towards 0 than if this regressors had been stationary. This effect will in fact dominate the influence of $(x_{2t}'u_{1t})$ for which a non-zero expected value would generally cause the OLS estimator of ξ to be biased and inconsistent because of simultaneity or missing regressors. The superconsistency of the OLS estimator comes from the fact that for any values of ξ that imply no cointegration between x_{1t} and x_{2t}, then $u_{1t} \sim I(1)$ so that the residual sum of squares explodes for large T. However, if the true value of ξ yields stationary residuals, the residual sum of squares will likely be much smaller for a large enough sample size. That is why any estimator aiming at minimizing some residual sum of squares will theoretically produce consistent values of a cointegrating relationship (if it exists) even though there may be a correlation between the $I(1)$ regressors and the error term.

---

[12] The normalization could have also been performed with respect to x_{2t}. Although this should not matter asymptotically, estimates of the cointegrating vector might be sensitive to the choice of the dependent variable in finite sample. This choice is sometimes based upon the theoretical preference of the researcher but the concept of cointegration remains atheoretical and does not favour one variable over the other. Because of the sometimes poor behaviour of the estimates of the elements of the cointegrating vector, in the individual equation context, the dependent variable may be chosen on the basis of the equation yielding the highest R^2.

[13] Technically, the rate of convergence of the OLS estimator for the coefficient of an $I(1)$ regressor is of order T or larger, while that for the coefficient of an $I(0)$ regressor is of order $T^{1/2}$.

Despite the theoretical superconsistency of the least-squares estimator of coefficients of $I(1)$ regressors, Phillips and Durlauf (1986) and Phillips (1991a) have showed that its distribution is not normally distributed. Its distribution is asymmetric as well as a function of nuisance parameters, which thus invalidates the use of standard distributions for inference about the cointegrating vector.[14]

Furthermore, with finite samples in practice, the theoretical superconsistency may not always be of great help. Monte Carlo experiments conducted by Banerjee *et al.* (1986) have shown that the number of observations may remain too small in many cases for the superconsistency of $\hat{\xi}$ to manifest itself.[15] Thus, a bias may contaminate the least-squares estimate and statistical tests may well falsely lead to the finding of no cointegrating relationships. This points out the need for developping better estimators than OLS, which will be discussed in further sections. In the bivariate case, as for equation (4.9), Banerjee *et al.* (1986) showed that this bias is inversely related to the R^2 of the regression.[16] Paquet (1992) gave an empirical illustration of the point raised by Banerjee *et al.* (1986) with actual data on realized and expected inflation. This example will be presented later.

4.3.3.1.2 Inference about cointegration in static regressions

To check empirically whether there is evidence of cointegration between time series, a natural way to proceed is to apply unit root tests on the residuals of a cointegration equation so that the null and alternative hypotheses are therefore:[17]

---

[14] The unit root in x_{2t} makes the distribution nonsymmetric, and this unit root along with endogeneity of x_{2t} makes the distribution dependent on nuisance parameters. Endogeneity of x_{2t} arises because of the long-run covariance between x_{2t} and u_{1t}. As discussed in Phillips and Loretan (1991) and Gonzalo (1994), strict exogeneity of x_{2t} makes the OLS on equation (4.13) asymptotically equivalent to the maximum likelihood estimation in a fully specified error correction model (MLECM) or Gaussian MLE. Hence, strict exogeneity of x_{2t} would allow the use of standard distributions to conduct inference on ξ_1.

[15] Furthermore, if x_{2t} is endogenous, the contemporanous correlation of Δx_{2t} with u_{1t}, or equivalently the long-run covariance between x_{2t} and u_{1t} brings about a simultaneous equation bias.

[16] A low R^2 should be seen as a warning sign. However, our practical experience suggests that there are cases, even in a bivariate setting with a fairly high R^2 for which evidence based on a static regression does not support the alternative cointegration hypothesis, while evidence based on dynamic regressions rejects the null of no cointregration.

[17] Park and Choi (1988), Park, Ouliaris, and Choi (1988), and Park (1990) have derived tests of the null hypothesis of

H_0: no cointegration <=> H_0: $u_{1t} \sim I(1)$
H_1: cointegration <=> H_1: $u_{1t} \sim I(0)$.

The residuals can be obtained from the OLS estimation of one of the following static cointegrating regression:

- the demeaned regression: $x_{1t} = \delta_0 + \xi x_{2t} + u_{1t},$ (4.15a)
- the demeaned regression with trend:

$$x_{1t} = \delta_0 + \delta_1 t + \xi x_{2t} + u_{1t} \ , \tag{4.15b}$$

- the demeaned regression with quadratic polynomial in time:

$$x_{1t} = \delta_0 + \delta_1 t + \delta_2 t^2 + \xi x_{2t} + u_{1t} \ , \tag{4.15c}$$

where both x_{1t} and x_{2t} are integrated, say of order one. These different specifications nest both alternative and null hypotheses. They can account for the alternative hypothesis of a cointegrating relationship, i.e. $[1, -\xi]\,[x_{1t}, x_{2t}]' \sim I(0)$, between two $I(1)$ variables when a long-run relationship exists either around a non-zero mean in (4.15a), or around a non-zero deterministic trend in (4.15b).[18] However, under the null hypothesis of no cointegration, empirical equations (4.15a), (4.15b) and (4.15c) imply that $[1, -\xi]\,[x_{1t}, x_{2t}]'$ is respectively a driftless nonzero mean integrated process, an integrated process with drift, or an integrated process with shifting drift. As discussed in Banerjee *et al.* (1993) and in Davidson and MacKinnon (1993), the inclusion of deterministic regressors (such as different power of time from 0 to 2) in the static cointegrating regressions is equivalent (at least asymptotically) to including them in the residual regressions used to perform the unit root test.

Residual based tests of no cointegration can then be applied as unit root tests on these residuals. If one estimates the following equation:

$$\Delta u_{1t} = (\rho - 1)\, u_{1t-1} + \sum_{i=1}^{l} \alpha_i \Delta u_{t-i} + v_t \quad . \tag{4.16}$$

An ADF test on u_{1t} can be applied along with the sequential data dependent procedure to select the lag order l discussed in the last chapter. An alternative consists in applying Phillips-Perron tests, $Z(t_\rho)$ and $Z(\rho)$, on the basis of the regression:

---

cointegration. Their procedure consists in testing first for a maintained order of a time polynomial then in testing the significance of extra superfluous time polynomial terms as a test of the null hypothesis of stationary residuals.

[18] In practice, the case where no constant would be needed in the cointegrating regression is rather seldom. Critical values of the non cointegration tests for this case are reported in Phillips and Ouliaris (1990).

$$\Delta u_{1t} = (\rho - 1) u_{1t-1} + v_t \,. \tag{4.17}$$

Even though the statistics for testing cointegration are constructed like unit root tests, there are important issues to take into account regarding their asymptotic distributions. First, because the cointegration tests are applied upon residuals obtained as a by-product of estimation rather than on raw data, the limiting distribution of the *ADF* and of various Phillips-Perron statistics is not that of unit root tests.[19] Second, as for the unit root tests on raw data, the distributions of the cointegration tests are affected by the set of deterministic regressors (e.g. intercept, linear time trend, squares time trend) in the cointegrating regression. Finally, the distribution depends also upon the number of $I(1)$ variables in the cointegrating regression (e.g. 2 in the examples considered in this section). Table A4.1 in the appendix of this chapter shows the asymptotic critical values of the *ADF,* as well of the two Phillips and Perron tests of no cointegration for different DGP (i.e. with an intercept only c, with a constant and a deterministic time trend cτ, and with a quadratic polynomial in time cττ). Notice that both the ADF statistics, denoted $t_{\rho i}$, and the Phillips-Perron $Z(t_{\rho i})$ test (for i = c, cτ, cττ) have the same asymptotic distribution. The limiting distribution of the non-parametrically corrected normalized bias test $Z(\rho_i)$ is different.

Phillips's and Ouliaris's (1990) Monte Carlo experiments suggest that the $Z(\rho_i)$ test is generally more powerful than the ADF or the $Z(t_{\rho i})$ tests but that their critical values are more sensitive to the parameters associated with the underlying DGP of the cointegrating residuals. On the other hand, the ADF and the $Z(t_{\rho i})$ statistics have more stable critical values even if they may be less powerful. Gregory's (1991) evaluation of the finite sample performance of various cointegration tests in the class of linear quadratic models leads to the recommendation of using ADF and Phillips-Perron $Z(\rho)$ tests for their seemingly reliable behaviour in terms of both test size and power.[20]

Moreover, as discussed earlier, theoretical superconsistency may not ensure the absence of finite sample biases in the estimation of a long-run relationship based on a static cointegrating regression. This could even prevent the rejection of the null hypothesis of no cointegration. It might therefore be preferable in many instances to consider jointly the long-run cointegrating relationships along with a statistical representation of the short-run dynamics.

---

[19] Only if one wanted to impose a value on the cointegrating vector, e.g. as suggested by economic theory, then the test of cointegration would be applied on a constructed linear combination of the original variables. It would then follow the same distribution as the unit root tests.

[20] These results are derived from a linear quadratic model in which economic agents make decision about variables of their choice by minimizing a multi-period quadratic cost function. The latter is defined as the sum of the cost occurring from the squared deviation about the long-run equilibrium value of the variable, and of the quadratic cost of adjustment. For instance, models of labour demand and supply, of money demand, and of computable general equilibrium have been derived in this set-up. See Gregory (1991) for references.

4.3.3.1.3 Cointegration and dynamic models

We began Chapter 4 discussing various equivalent dynamic specifications that were shown to be consistent with an existing long-run relationship between $I(0)$ variables. Since it has been shown in the last section that cointegrating relationships between $I(1)$ variables are in effect long-run relationships, it appears natural to extend dynamic specifications to the case of $I(1)$ time series.

Indeed, Engle and Granger (1987) have established that there exists a corresponding ECM representation associated with a cointegrating relationship. For instance, suppose that the variables y_t and x_t are individually integrated, that they are cointegrated, and that x_t is weakly exogenous with respect to ξ_1 in the following cointegrating equation:

$$y_t = \xi_0 + \xi_1 x_t + u_t \ . \tag{4.18}$$

It follows that a corresponding ECM is given by:[21]

$$\Delta y_t = \kappa + \sum_{i=1}^{p-1} \varphi_{i1} \Delta y_{t-i} + \sum_{i=0}^{p-1} \varphi_{i2} \Delta x_{t-i} + \alpha\,(u_{t-1}) + \varepsilon_t \,. \tag{4.19}$$

We can call ξ_1 the long-run multiplier, while φ_{02} is a short-run or impact multiplier.

Relying on the superconsistency of the coefficients of the $I(1)$ regressor in the static cointegrating regression, Engle and Granger (1987) have suggested the following two steps procedure:

1. Having checked that one is dealing with integrated time-series, first estimate a static cointegration equation and test for evidence against the null hypothesis of no cointegration. This could be evidenced by residual based tests on equation (4.18). If there is such evidence, go on. If not, go back to the drawing board.

---

[21] In this individual equation specification, as well as in those that follow, the introduction of an unrestricted constant is intended to capture, when applicable, both a drift in the dynamics and a mean effect in the long-run relationship. This is why if the error correction term is written explicitly, care should be taken not to include two separate intercepts in the empirical equation.

2. If there appears to be cointegration, use the first lag of the residuals from the relevant static cointegrating regression, $\hat{u}_t$, as the estimate of the error-correction term in an ECM representation, such as (4.19).

One problem with Engle's and Granger's (1987) two steps procedure is that the least-square estimates of the static cointegrating regression may suffer from significant biases in finite samples. In this case, another possibility would be to estimate directly the ECM equation,

$$\Delta y_t = \kappa + \sum_{i=1}^{p-1} \varphi_{i1} \Delta y_{t-i} + \sum_{i=0}^{p-1} \varphi_{i2} \Delta x_{t-i} + \alpha\,(y_{t-1} - \xi_1 x_{t-1}) + \varepsilon_t\,, \tag{4.20}$$

in one step by non-linear least-squares. The estimates of α and ξ_1 will then be more satisfactory than those reckoned by the two-step procedure, and the computer output will provide direct estimates of their standard error. As discussed by Stock (1987) and Phillips (1991a), the application of NLS does not generally guarantee that the estimator of ξ_1 is asymptotically standard normal, as its distribution remains dependent of nuisance parameters. Therefore, the standard error at most gives some idea of the precision of the ξ_1 estimate, the NLS slope estimate should not be used for inference with usual distribution tables.[22] Finally, if one is only interested in the coefficient values, it is easy to distribute the α coefficient on each term of the error-correction term in equation (4.20), and to estimate the resulting equation by OLS. The NLS estimate of ξ_1 can then be recovered as minus the ratio of the coefficient estimate of x_{t-1} on that of y_{t-1}. Notice also that residual based cointegration tests can be conducted on either $\left[y_{t-1} - \hat{\xi}_1 x_{t-1}\right]$ or ε_t.

As discussed in section 4.2, yet another possibility is to estimate directly the long-run multiplier (i.e. the coefficients of the cointegrating vector since we are dealing with $I(1)$ variables) in a Bewley representation of the dynamic model. In this case, the application of the GIVE procedure on an equation like

$$y_t = \xi_0 + \xi_1 x_t + \sum_{i=0}^{m} \eta_i \Delta y_{t-i} + \sum_{j=1}^{n} \psi_j \Delta x_{t-j} + v_t \tag{4.21}$$

---

[22] Gonzalo (1994) explains that full system MLE of ξ_1, which would be standard normal, is asymptotically equivalent to NLS on equation (4.20) if the impact multiplier φ_{02} is zero. Notice that the value of φ_{02} has no bearings on the weak or strong exogeneity of x_t with respect to ξ_1.

will generally provide consistent estimates of the long-run multiplier. As such, equation (4.21) looks like a dynamic generalization of the static cointegrating equation. Lagged first-differences in both y_t and x_t are introduced to capture explicitly the dynamics that would have been left otherwise in the error term. The ADF and the Perron-Phillips residual based tests of no cointegration can also be applied on the residuals of equation (4.21). Nonetheless, Phillips and Loretan (1991) showed that, unless x_t was strictly exogenous, the distribution of the GIVE estimator of ξ_1 remains non-standard since it is nonsymmetric and nuisance parameters dependent.

As we discussed earlier, we would generally require that x_t and the Δx_{t-j}'s be weakly exogenous with respect to the parameters of interests for the estimation of the long- and short-run coefficients to be correctly estimated. The $I(1)$ variable x_t would fail to be weakly exogenous, if there exists a dynamic representation for x_t, such as

$$\Delta x_t = \kappa + \sum_{i=1}^{p-1} \varphi_{xi2}\,\Delta x_{t-i} + \sum_{i=0}^{p-1} \varphi_{yi2}\,\Delta y_{t-i} + \alpha_x\,(y_{t-1} - \xi_1 x_{t-1}) + \varepsilon_{xt}\,, \tag{4.22}$$

with $\alpha_x \neq 0$, i.e. where the ECM term, or lagged deviation from the long-run equilibrium, appears in the dynamic equations for both Δy_t and Δx_t.[23, 24] In this case, because x_t is $I(1)$, the estimator of ξ_1 is superconsistent and will be consistently estimated, despite the endogeneity of x_t. This will not be true for the coefficients associated with the short-run dynamics because it is described by first-differenced variables that are $I(0)$. Furthermore, as explained in Banerjee *et al.* (1993), the lack of weak exogeneity will have adverse effects on finite sample biases and on efficiency in the separate estimation of the long-run multiplier from the individual equations defined as (4.18), (4.20) or (4.21). At the same, as we have previously indicated, the distribution of these estimators is typically non standard. We discuss next alternative estimation methods that are asymptotically optimal, i.e. consistent, efficient and that are equivalent to Gaussian MLE, for an individual equation with endogenous $I(1)$ regressors.

---

[23] This suggests that evidence of weak exogeneity for $I(1)$ variables in a given ECM representation can be found from a joint estimation of the ECM representations of all variables, and from testing for the significance of the error-correction terms (or of cointegrating vectors) in the other equations of the system. This can be done using Johansen's (1988) FIML method.

[24] A necessary condition for weak endogeneity of x_t with respect to ξ_1 requires a zero coefficient of the error correction term in an ECM like (4.20) with Δx_t as dependent variable. As shown by Gonzalo (1994), if in addition $\varphi_{x0} = \xi_1$ in equation (4.20), then x_t is strictly exogenous with respect to ξ_1 in a regression of y_t on x_t.

4.3.3.1.4 Optimal estimators of cointegrating coefficients with asymptotic standard distribution in individual equations with endogenous regressors.

When $I(1)$ regressors are endogenous, several single equation estimation methods can be used. These are all alternatives to the joint estimation of the system of dynamic equations, suggested by Johansen (1988), that are characterized by desirable asymptotic properties with respect to efficiency and standard distributions like Johansen's (1988) approach. This latter approach will be the object of a later section. We now focus on single dynamic equation estimation.

Phillips and Hansen (1990) have proposed a fully modified least-squares estimator to estimate the static cointegrating regression, e.g. equation (4.18). Starting with the OLS estimates of the cointegrating regression, they apply to it two nonparametric corrections that make the modified estimator follow the same asymptotic distribution to that of a maximum likelihood estimator applied to a system of dynamic error-correction equations for each endogenous regressors (MLECM). The first correction takes account of any serial correlation in the residuals, while the second modification corrects for the endogenous influence of y_t on x_t using a normalized nonparametric estimate of the long-run covariance between the respective innovations on y_t and x_t. Also, a modified t-statistic to test a null hypothesis about ξ_1 asymptotically follows the standard normal distribution. Simulations by Phillips and Loretan (1991) show that the fully modified estimator may behave poorly in finite samples when the error term of the model exhibits negative serial correlation.[25]

Phillips (1991c) and Phillips and Loretan (1991) have also proposed a spectral regression method that casts the cointegrating regression in the frequency domain rather than in the time domain. This estimator also enjoys the asymptotic properties of the full system MLE.[26] Readers interested in constructing the fully modified or the spectral regression estimator should refer to the cited papers.

---

[25] This is analogous to the problem of Phillips-Perron unit root tests, with negative MA.

[26] Essentially, it consists in applying a band spectral estimation on a single linear equation. The dependent variable is the discrete Fourier transforms of y_t and the regressors are the discrete Fourier transforms of x_t and Δx_t. Estimation is then carried out by feasible GLS, using estimates of the spectra of v_t as weights.

Saikkonen (1991) has proposed instead to deal with the endogeneity problem with a parametric correction to the least-squares estimator of the cointegrating vector. This amounts to estimate the following equation in place of the static cointegrating regression:

$$y_t = \xi_0 + \xi_1 x_{1t} + \sum_{i=1}^{P} f_{1i}\, \Delta x_{1t-i} + \sum_{j=1}^{Q} f_{2j}\, \Delta x_{1t+j} + v_t \ . \tag{4.23}$$

Lags and leads of the first difference in x_t are added in the static cointegrating equation. Even though the orders, P and Q, respectively of the lag and lead polynomials could theoretically be infinite, significance tests can be applied to determine the parsimonious adequate representation down from a general model. Intuitively, this procedure relies upon the idea that the endogeneity of x_{1t} means that the innovations on y_t Granger cause the innovation on x_{1t}. Hence, future values of the innovations on x_{1t} may contain information helpful in predicting y_t, which explains the role of the lead values of Δx_{1t} in the regression. To correct for serial correlation left in the error term, Stock and Watson (1993) propose either to adjust non-parametrically the variance-covariance matrix and the relevant Wald statistics, or to apply generalized least-squares on equation (4.23). These are called the dynamic OLS and the dynamic GLS.

Phillips and Loretan (1991) build on equation (4.23) but suggest to deal parametrically for serial correlation, by adding lags of the error-correction term up to a sufficient order for the modified equation's error term to be white noise. Because of the unit roots induced persistence effects of the innovations, these extra regressors generally constitute a more adequate proxy than lagged Δy_t's for the past history of u_t in equation (4.18). The significance of extra lags of the error-correction term can be tested with conventional statistical methods. This may be particularly important because their simulations suggest that an overfitting of the lag length on this term can make sensitive the results on inference about the cointegrating vector. Then, the modified equation, i.e.

$$y_t = \xi_0 + \xi_1 x_{1t} + \sum_{i=1}^{P} f_{1i}\, \Delta x_{1t-i} + \sum_{j=1}^{Q} f_{2j}\, \Delta x_{1t+j} + \sum_{k=1}^{R} f_{3k}\,[y_{t-k} - \xi_1 x_{1t-k}] + \omega_t \ , \tag{4.24}$$

can be estimated by non-linear least squares. This procedure is asymptotically efficient, asymptotically median unbiased, and justifies the use of conventional chi-squared type criteria for inference on the long-run coefficients.

4.3.3.2 Generalisation to the N variables case.

We have discussed so far econometric issues related to cointegration, the specification of dynamic models consistent with cointegrating relationship, as well as how to get correct estimates in different context, building up our understanding around cointegration between two $I(1)$ variables. There are many instances for which a cointegrating relationship may involve more than two variables (e.g. a long-run money demand equation as in Ambler and Paquet (1992)). All the issues and methods discussed so far can be straightforwardly extended to a cointegrating relationship between N variables.

Some complications arise however in the case of N integrated variables. First, if some $I(1)$ variables have been wrongly omitted from the equation used to look for cointegration, the residuals may be integrated. It is therefore always advisable to begin with the more general model. Second, with possibly N variables entering the cointegrating relationship, we face many alternative normalisations leading to the choice of the dependent variable in either a static cointegrating equation, or in a dynamic representation. Note that this normalisation makes sense only by imposing *a priori* that the cointegration coefficient associated with the chosen dependent variable is non-zero. Third, as we mentioned in section 4.3.2, with $N > 2$ integrated variables, there may exist up to $N-1$ cointegrating vectors. Insofar as there exists more than one cointegrating relationship, a particular normalization and corresponding static or dynamic representation can therefore be thought only as providing consistent estimates of a vector lying in the cointegrating space. Another way to put it is that the possibility of non-uniqueness of the cointegrating vectors may imply that more than one error-correction term enters the dynamic specification. But, the multiplicity of cointegrating relationships leaves them unidentified unless we impose identifying restrictions. Phillips (1991b), explained how this is analogous to the issues of identification in a standard simultaneous equations model. Identifying restrictions are required so that one can uniquely determined a structural form from a given reduced form.

An example of non uniqueness is that if N integrated variables are cointegrated, it is possible that some subsets of these variables be cointegrated as well. Indeed, linear combinations of cointegration relationships will also be stationary. That is why, when working with more than two integrated regressors, it may always be preferable to investigate all the possible combinations that may reflect a cointegrating relationship, at the stage of characterising the data. This is particularly important in the light of Choi 's (1994) recent findings that the conventional critical values (such as those reported in Table A4.1) for residuals based tests of non cointegration are not appropriate when some of the regressors are cointegrated. In fact, the true asymptotic distributions of these tests are function of nuisance parameters involving generally unknown information with respect to the cointegrating structure of the regressors.

With *N* integrated variables, the optimal single-equation procedures remain applicable when there is only one underlying cointegrating relationship. Nonetheless, Phillips and Lorethan (1991) explain that the optimal single-equation procedures keep their desirable asymptotic properties even if there is more than one cointegrating relationship, "provided that there are no across-relationships restrictions on the parameters of the long-run equilibria." (p.416) However, since no *a priori* restrictions (but the normalization *via* the choice of the dependent variable) are imposed in the single-equation procedures, the estimates hence obtained do not identify any individual structural cointegrating equation. As reported in Hamilton (1994), Woolridge (1991) has shown that optimal single-equation estimation procedures will typically select a cointegrating relation whose residuals are uncorrelated with any other $I(1)$ linear combinations of the RHS regressors. This estimated relation may yet be quite different from the structural cointegration relationship even though it also lies in the cointegrating space.

When there are say $r > 1$ cointegrating vectors, Campbell and Perron (1991) have proposed to use the single-equation methods on a stacked system of r equations. Such a system can be estimated by multivariate (linear or nonlinear) least-squares. In any case, one must still assess how many cointegrating vectors there are amongst the *N* integrated variables. A maximum likelihood system estimation method proposed by Johansen (1988) deals with the endogeneity of the regressors, and can be used to uncover the dimension of the cointegrating space, to estimate a basis for that space, and to test restrictions on the cointegrating relationships. This will be addressed in section 4.4.

4.3.3.3 Inference about parameters of integrated regressors

In an econometric model involving $I(0)$ regressors, references to basic central limit theorems ensure that the least-squares estimator of the coefficient is asymptotically normally distributed. However, Park and Phillips (1988, 1989) have shown that the asymptotic distribution of the least-squares estimators of coefficients of $I(1)$ regressors is not necessarily standard normal.[27] That is why the usual t-, F- or χ^2-tests are not applicable as is, except in particular circumstances. Unless we can rely on proven theorems that validate the use of asymptotic standard distributions, the remaining possibility is to proceed with Monte-Carlo integrations. These simulations can generate the finite sample distribution of the estimator and the correct critical values of the tests on coefficients of integrated regressors, as a function of the sample and the underlying generating process of the data at hands.[28]

Fortunately, there are particular instances for which Monte-Carlo experiments are not required, if one is willing to rely on asymptotic results. Below is a list of many important results found in the literature concerning equations estimated by OLS, GIVE, or NLS.[29]

1. The least-squares estimators of coefficients of stationary regressors always follow asymptotically a normal distribution whether there is an $I(1)$ dependent variable, or whether there are other integrated regressors in the empirical equation.
2. If an $I(1)$ regressor is strongly exogenous, the least-squares estimator of its coefficient is asymptotically normal.
3. If there is only one $I(1)$ regressor in the equation, if its DGP includes a nonzero drift, and if there is no deterministic trend included as a regressor, then the least-squares estimator of its coefficient is asymptotically normal.[30]
4. Because of the presence of $I(1)$ regressors, the coefficients on the intercept or any polynomials in time included in the regression and their associated t statistics typically have a nonstandard distribution.
5. Test statistics of joint hypotheses on any subset of the set of $I(1)$ variables do not follow an asymptotic standard distribution. Moreover, test statistics on joint hypotheses involving any

---

[27] In a particular instance, this had already been shown by Fuller (1976) and by Dickey and Fuller (1981) when dealing with the t-test of a unitary first-order autoregressive coefficient, namely regarding the distribution of unit root tests.

[28] The description of the steps and methods involved to perform correctly Monte-Carlo experiments goes beyond the scope of this guide. Interested readers can refer to Davidson and MacKinnon (1993) who provide a very good chapter on this topic. See also Pagan and Wickens (1989) and Watson (1993).

[29] See Park and Phillips (1988, 1989), West (1988a), Sims, Stock and Watson (1990), and Watson (1993).

[30] If there are more than one $I(1)$ regressor in the equation, only a very specific linear combination of the coefficients of the two $I(1)$ variable would have a LS estimator that is asymptotically normal. Hence, this result does not imply that tests on these coefficients would follow a standard distribution. Also, if a deterministic trend appears as well amongst the regressors, this is analogue to have detrended the drifted $I(1)$ regressor. This removes the role of the drift that led to a standard distribution of the LS estimator for the coefficient of the single integrated regressor.

stationary regressors in combination with any $I(1)$ regressors do not either follow an asymptotic standard distribution.

6. In a regression with many $I(1)$ regressors, if there exists at least one transformation of the regression equation that casts the hypotheses of interest as restrictions on zero-mean $I(0)$ regressors, the test statistics of these hypotheses based on the estimates of the original regression follow a standard distribution. This applies to individual t-tests, as well as to tests on a joint hypothesis on the proper subset of $I(1)$ regressors.[31]
7. If a reparameterization of the original regression can be rewritten using cointegration between some of the $I(1)$ regressors in a way that the transformed regression is expressed in terms of additional $I(0)$ regressors, then joint test statistics involving these $I(0)$ regressors with any of the other $I(0)$ regressors follow standard distributions asymptotically.
8. Test statistics constructed for testing restrictions on linear combinations of coefficients of zero mean $I(0)$ regressors together with coefficients of $I(1)$ regressors and/or with deterministic trends follow asymptotic standard distributions.

The first four propositions are quite straightforward to visualize, the implications of propositions 5, 6, 7 and 8 in applied work will be best understood with some illustrations.

For instance, consider three time-series, y_t and x_t that are $I(1)$, and $z_t \sim I(0)$, but that the long-run relationship between the two $I(1)$ variables is given by:

$$y_t = \xi_0 + x_t, \tag{4.25}$$

so that y_t and x_t are $CI(1,1)$. Suppose that one estimates the following empirical model by OLS,

$$y_t = \tau_0 + \tau_{y1} y_{t-1} + \tau_{x1} x_{t-1} + \tau_{x2} x_{t-2} + \tau_{x3} x_{t-3} + \tau_z z_t + u_t. \tag{4.26}$$

Suppose that the researcher considered testing the different following null hypotheses with some Wald- (F-, or, when applicable, t-) statistics:

$$H_{A0}: \ \tau_{y1} = 1, \ \tau_{x1} = \tau_{x2} = \tau_{x3} = 0 \ ; \tag{4.27a}$$

$$H_{B0}: \tau_{y1} = 1; \tag{4.27b}$$

$$H_{C0}: \tau_{xi} = 0, \text{ for some } i = 1,2, \text{or } 3 \ ; \tag{4.27c}$$

$$H_{D0}: \tau_{x1} = \tau_{x2} = \tau_{x3} = 0 \ . \tag{4.27d}$$

$$H_{E0}: \tau_{x1} + \tau_z = 0 \ ; \tag{4.27e}$$

---

[31] The presence of an intercept in the regression ensures that the $I(0)$ regressors are demeaned.

By adding and subtracting $[\tau_{y1} x_{t-1}]$ to the RHS of (4.26), we get the following reparameterization:

$$y_t = \tau_0 + \tau_{y1}(y_{t-1} - x_{t-1}) + (\tau_{y1} + \tau_{x1})\, x_{t-1} + \tau_{x2} x_{t-2} + \tau_{x3} x_{t-3} + \tau_z z_t + u_t \ , \qquad (4.28a)$$

so that τ_{y1} can be expressed as the coefficient of a stationary regressor. Also, by adding and subtracting $[\tau_{x1} y_{t-1} + \tau_{x2} y_{t-2} + \tau_{x3} y_{t-3}]$ to the RHS of (4.26), there exists a reparameterization expressing τ_{x1}, τ_{x2}, and τ_{x3} as coefficients of zero mean $I(0)$ regressors, as seen from:

$$\begin{aligned} y_t = {} & [\tau_0 + \tau_{x1}\xi_0 + \tau_{x2}\xi_0 + \tau_{x3}\xi_0] + [\tau_{y1} + \tau_{x1} + \tau_{x2} + \tau_{x3}]\, y_{t-1} \\ & + \tau_{x1}[x_{t-1} - y_{t-1} - \xi_0] + \tau_{x2}[x_{t-2} - y_{t-1} - \xi_0] + \tau_{x3}[x_{t-3} - y_{t-1} - \xi_0] + \tau_z z_t + u_t \ . \qquad (4.28b) \end{aligned}$$

From proposition 5, a Wald test of H_{A0} will not follow a standard asymptotic distribution. However, because of the existence of representations (4.28a) and (4.28b), proposition 6 implies that a Wald test of either H_{B0}, H_{C0}, or H_{D0} can be conducted with reference to the usual critical values. Finally, proposition 8 implies also that H_{E0} can be tested with statistics that follow standard asymptotic distributions.

Still considering the same three time-series, suppose now that the researcher estimates the following model by OLS:

$$\Delta y_t = \delta_0 + \delta_1 t + \delta_{y1} y_{t-1} + \delta_{y2} y_{t-2} + \delta_{x1} x_{t-1} + \delta_z z_t + v_t \ . \qquad (4.29)$$

Here again, individual *t*-tests on any coefficients of equation (4.29) would asymptotically follow a standard normal distribution, because of proposition 6, although a joint test on δ_{y1} and δ_{x1} should not be performed using the critical values extracted from the usual tables. Meanwhile, if one were interested in estimating, say the following transformation of equation (4.29):

$$\begin{aligned} \Delta y_t &= \delta_0 + \delta_1 t + \delta_{y1}(y_{t-1} - x_{t-1}) + \delta_{y2}(y_{t-2} - x_{t-1}) + (\delta_{y1} + \delta_{x1} + \delta_{y2}) x_{t-1} + \delta_z z_t + v_t \ , \\ \Delta y_t &= \delta_0 + \delta_1 t + \delta_{y1}(y_{t-1} - x_{t-1}) + \delta_{y2}(y_{t-2} - x_{t-1}) + \eta x_{t-1} + \delta_z z_t + v_t \ . \qquad (4.30) \end{aligned}$$

In this case, test statistics of a joint null hypothesis regarding say δ_{y1}, δ_{y2} and δ_z would follow standard distribution (proposition 7). But, test statistics of a joint null hypothesis involving η and any other stationary regressors will require that one use computer simulations to compute the appropriate critical values (proposition 5).

4.3.3.4 How to deal with I(1) regressors in regressions: the case of the permanent income theory of consumption

Stock and Watson (1988) have constructed an example (cited also in Banerjee *et al.* (1993)) that illustrates many of the pitfalls to be avoided and the care one must exercise in dealing with integrated regressors and regressand. We build on this example below to review some important issues. Let us suppose that we have data expressed in logs on real nondurable consumption, c_t, on real disposable income, y_t, and on the price level, p_t. Let us further assume that the underlying data generating process is based on the permanent income theory of consumption according to the following equations:

$$y_t = y_t^P + y_t^S \,, \tag{4.31}$$

$$y_t^P = \tau + y_{t-1}^P + u_t \,, \tag{4.32}$$

$$c_t = y_t^P \,, \tag{4.33}$$

$$p_t = p_{t-1} + v_t \,, \tag{4.34}$$

where y_t^P and y_t^S are respectively the permanent and the transitory components of real disposable income, the latter being a stationary innovation on y_t. The parameter τ is greater than or equal to zero, and $E(u_t\, v_t)$ is assumed to be zero.

According to the permanent income theory of consumption, rational economic agents who maximize their lifetime utility with respect to their intertemporal budget constraint over the same horizon will seek to smooth out their consumption over time. Hall (1978) has shown that with time-separable-quadratic preferences, real consumption would then behave as a random walk (with or without drift). The DGP summarized by equations (4.31) to (4.34) is compatible with this theory.

4.3.3.4.1 Application 1

Assume that a researcher wants to test for the existence of monetary illusion in economic agents' consumption decisions. Contrary to a relative price, the price level on its own should provide no information regarding the scarcity of resources and should not affect real consumption. He then proceeds in estimating the following equation by OLS and in testing H_0: $\beta_1 = 0$ versus H_0: $\beta_1 \neq 0$.

$$c_t = \alpha_1 + \beta_1 p_t + e_{1t} \tag{4.35}$$

According to their DGP, since c_t and p_t are two independent integrated processes, equation (4.35) is a spurious regression so that e_{1t} should be $I(1)$ and there is no cointegration between the two variables. If the researcher were not aware and were not testing for this problem, he would most likely be led to reject wrongly the null hypothesis as the t-statistic would tend to be high. Moreover, remember that the OLS estimator of β_1 converges to a random variable and that it is not generally distributed normally in the case of an $I(1)$ regressor, so the t-statistic would not follow the usual distribution. This remains true especially since p_t is an $I(1)$ with no drift. Anyhow, as emphasized by Banerjee *et al.* (1993), "inference of this kind would appear to be faulty, in any case. To consider a rejection of H_0: $\beta_1 = 0$ as a reason for accepting any specific alternative is statistically unjustifiable."

4.3.3.4.2 Application 2

Suppose that the researcher wants to check whether real consumption has a deterministic trend on the basis of the estimation of:

$$c_t = \alpha_2 + \beta_2 t + e_{2t} \ . \tag{4.36}$$

Here again, this is a spurious relationship since c_t is a difference stationary time series, not a trend stationary one. The error term e_{2t} should be $I(1)$ as this is not a cointegrating relationship. A deterministic trend does not remove the contribution of a stochastic trend in a series.

4.3.3.4.3 Application 3

Suppose that the researcher wants to estimate the marginal propensity to consume from the following regression by OLS:

$$\Delta c_t = \alpha_3 + \beta_3 \Delta y_t + e_{3t} \ . \tag{4.37}$$

This is a balanced regression since both the regressand and the regressor are $I(0)$. In fact, the researcher is following the prescriptions of Box-Jenkins that one can work with the first-differenced series, when the variables in level are integrated. Unfortunately, this specification is suffering from the omitted regressor problem since it fails to take into account that there is a long-run cointegrating relationship between c_t and y_t. Therefore, the OLS estimator of β_3 is likely to be biased and inconsistent since the change in real disposable income is the sum of the changes in both permanent and transitory real disposable income and e_{3t} is correlated with Δy_t. Consequently, it would have been preferable to test for a cointegrating relationship between c_t and y_t and to estimate an error-correction model, such as:

$$\Delta c_t = \alpha_3 + \beta_3 \Delta y_r + \gamma_3 [y_{t-1} - c_{t-1}] + \tilde{e}_{3t}. \tag{4.38}$$

Moreover, because of the possibility that y^s_t and therefore e_{3t} be serially correlated, the dynamics could be richer in (4.38) with the inclusion of lagged Δc_t and Δy_t.

4.3.3.4.4 Application 4

The researcher estimates the following equation by OLS and tests H_0: $\beta_4 = 0$:

$$\Delta c_t = \alpha_4 + \beta_4 y_{t-1} + e_{4t} . \tag{4.39}$$

Since Δc_t~$I(0)$ and y_{t-1}~$I(1)$, therefore e_{4t}~$I(1)$. This is an unbalanced regression so that the *t*-statistic is inflated and leads too easily to the rejection of the null hypothesis, and it fails to take into account that the distribution of the estimator of β_4 is not necessarily normal since it is associated with an integrated regressor. Notice however that since there is only one $I(1)$ regressor, if there is a drift in the DGP of this variable, i.e. $\tau \neq 0$, then the OLS estimator of β_4 is asymptotically normal.

4.3.3.4.5 Application 5

Instead, suppose the researcher estimates instead the following equation:

$$c_t = \alpha_5 + \beta_5 y_t + e_{5t} \tag{4.40}$$

before testing H_0: $\beta_5 = 0$. In this case, this amounts to a static cointegrating regression, e_{5t}~$I(0)$, which can be tested. The OLS estimation of this balanced equation will also provide in theory a superconsistent estimator of β_5. Care should be taken however for the possibility that in finite sample there may remain a substantial bias since much dynamics could be left in the residuals. Furthermore, generally the OLS estimator of β_5 would not be asymptotically normally distributed unless $\tau \neq 0$ in this case with a single integrated residual. In addition, serial correlation in the error term makes the usual OLS estimator of the standard error incorrect. This would therefore suggest the construction of an HSCC covariance matrix.

For these reasons, it might be preferable to model statistically in an explicit manner some of the dynamics by adding in lags of the first-difference of both the regressand and the regressor, even though the estimation should then be carried out by GIVE.

4.3.3.4.6 Application 6

The researcher decides finally to estimate:

$$c_t = \alpha_6 + \beta_6 c_{t-1} + \gamma_6 y_{t-1} + e_{6t} \ . \tag{4.41}$$

This is also a balanced equation and $e_{6t} \sim I(0)$, since there exists long-run equilibrium between c_t and y_t, i.e. $c^* = (\alpha_6/1-\beta_6) + (\gamma_6/1-\beta_6)\, y^*$ under cointegration. Since two $I(1)$ regressors appear in the empirical equation, one would expect the OLS estimator of β_6 or of γ_6 not to follow the asymptotically normal distribution, notwithstanding the presence of a drift in the DGP of both c_t and y_t. However, as shown by Sims, Stock and Watson (1990), since there exists reparameterizations of equation (4.41) that express the coefficients β_6 and γ_6 as coefficients of $I(0)$ regressors, t-statistics constructed to test single hypothesis on either of these coefficients follow asymptotically the normal distribution. These reparameterizations can be derived by adding and subtracting $\beta_6 y_{t-1}$ or $\gamma_6 y_{t-1}$ on the RHS of equation (4.41), thus yielding:

$$c_t = \alpha_6 + \beta_6 [c_{t-1} - y_{t-1}] + (\beta_6 + \gamma_6)\, y_{t-1} + e_{6t} \ , \tag{4.42a}$$

$$c_t = \alpha_6 + (\beta_6 + \gamma_6)\, c_{t-1} + \gamma_6 [c_{t-1} - y_{t-1}] + e_{6t} \ . \tag{4.42b}$$

However, an F-statistic reckoned to perform a joint test of a null hypothesis about both β_6 and γ_6 would have a non-standard distribution since any reparameterization of equation (4.41) would involve an $I(1)$ and an $I(0)$ regressor.

The different cases considered in this subsection illustrate many of the issues one should worry about when working with dynamic regressions that include integrated regressors. In this example, as argued by Banerjee *et al.* (1993), "a general to specific method of econometric modelling would have overcome many of the problems of spurious inferences and non-standard distribution." (p. 192). Thus, the consideration of a dynamic model between several regressors can be often advisable. However, notice that this does not preclude the need for the regressors to be weakly exogenous for the parameters to be estimated consistently and efficiently. If it is not the case, the discussions of sections 4.3.3.1.3 and 4.3.3.1.4 need to be taken into account.

4.3.4 Example: inflationary expectations and rationality I

Paquet (1992) investigated whether deviations between survey measurements of inflationary expectations and realized inflation tend to persist over time. Cukierman and Meltzer (1982) had shown that following a large permanent supply shock, the possibility of confusion about the duration of the shock (i.e. whether it is transitory or permanent) can account for serial dependence of the forecast errors in finite samples without implying violation of the rational expectations hypothesis (REH). Provided that inflationary expectation and realized inflation are both integrated, under rationality the long-run response of expected inflation to the actual inflation rate should equal one. This means that economic agents should not be systematically fooled, and forecasting errors should not persist over time.

The sample of data consists of semi-annual observations of the six-month-ahead expected actual rates of change of the U.S. CPI from the second half of 1951 to the first half of 1990. The actual inflation rate was calculated from Citibase and the expected inflation series is that of the Livingston survey.

Table 4.4 reports results from unit root tests conducted on both series. Because of the usual low power of unit root tests, it is advisable to confront the series to many tests. The null hypothesis of a unit root is not rejected even at 5% for the two series according to the ADF test,

Table 4.4- Unit root tests for actual and Livingston measure of expected inflation rate.
Sample: 1951H2 - 1990H1 (T=77).[1,2]

Series	Lag truncation parameter value	PI_t	PI^e_t	Null hypothesis
$\hat{\rho}$		0.9153	0.9877	
t_{μ}	$l=3$	-2.48	-1.56	$I(1)$
	$l=11$	-1.34	-1.29	$I(1)$
$Z(\hat{\rho}_{\mu})$	$l=3$	-16.54*	-3.23	$I(1)$
	$l=11$	-19.11*	-4.71	$I(1)$
$Z(t_{\mu})$	$l=3$	-3.28*	-1.31	$I(1)$
	$l=11$	-3.08*	-1.55	$I(1)$
$Z(\Phi_1)$	$l=3$	5.41***	0.90	$I(1)$ and nonzero mean
	$l=11$	4.79***	1.23	$I(1)$ and nonzero mean

[1] $\hat{\rho}$ is the estimate of the first-order autocorrelation in $y_t = \rho y_{t-1} + u_t$.

[2] Rejection level of H_0 on the basis of Fuller (1976) and Dickey and Fuller (1981) critical values: * at 1%; ** at 2.5% *** at 5%.

$t_{\rho c}$. At the 2.5% level, the evidence is against the presence of a unit root in PI_t according to the Phillips-Perron tests while the survey measure of expected inflation appears to be integrated. Application of the tests is, however, not straightforward if the first difference of the actual inflation series exhibits negative moving average serial correlation. As shown by Schwert (1987), for an $IMA(1,1)$ process with a negative moving average coefficient, the usual critical values for the Phillips-Perron tests will reject too often the unit root hypothesis. However, the ADF test might suffer less from this problem. Estimating an $IMA(1,1)$ process for the PI_t series yielded a statistically significant (at the 1% level) MA coefficient equal to -0.38. Interpolating from the critical values tabulated in Schwert (1987) for both the $Z(\hat{\rho}_c)$ and the $Z(t_{\rho c})$ statistics, their calculated values fall in the non-rejection region of integration.[32,33] Overall, the evidence is consistent with the presence of a unit root in the two series. The next step consists in testing whether both series are cointegrated in a way that forecast errors do not persist.

The first column of Table 4.5 shows the estimates and the cointegration tests for the demeaned static cointegrating regression. While both Phillips-Perron tests reject the null hypothesis of no cointegration at the 1% level, the augmented Dickey-Fuller test rejects it only at the 10% with $l=3$. Although the OLS estimator of β_1 is theoretically superconsistent under cointegration, the low R^2 of this regression (0.6661) suggests some caution in interpreting an OLS estimate of β_1 equal to 0.6526. As shown by the Monte Carlo experiments of Banerjee *et al.* (1986) there may well be a substantial bias in finite samples. That is why Paquet (1992) then proceeded in estimating one of Bewley transformations of a dynamic equation relating expected and realized inflation:

$$PI^e_t = \beta_0 + \beta_1 PI_t + \Sigma_{i=1}^{8} \eta_i \, \Delta PI^e_{t-i} + \Sigma_{j=1}^{8} \psi_j \, \Delta PI_{t-j} + u_t \, . \tag{4.43}$$

The OLS estimates of β_0 and β_1 are shown for illustrative purposes in column #2 of Table 4.5. But since the variable $\Delta PI^e_t = PI^e_t - PI^e_{t-1}$ is included amongst the regressors and is thus correlated with the error term, this equation should be estimated by GIVE. As shown in column #3, the cointegration tests based on the residuals of this equation support the existence of

---

[32] For instance, from his simulations for an IMA(1,1) process with an MA coefficient equal to -0.5 rather than 0, with 50 observations and a lag length $l=3$, Schwert (1987) reports 5% critical values of -35.1 instead of -13.8 for the $Z(\rho_c)$ test, and of -5.14 instead of -2.97 for the $Z(t_{\rho c})$ test. (See Table A3.7 and A3.8).

[33] Paquet (1992) also reported results from Park and Choi's (1988) tests rejecting the null hypothesis of stationarity for both PI_t and PI^e_t at the 1% significance level.

Table 4.5- Cointegration tests between PI_t and PI_t^* from efficient modified estimators.
(Sample: 1951H2 - 19990H1).[1]

Equation: $PI_t^* = \beta_0 + \beta_1 PI_t + \sum_{i=0}^{r} \eta \, \Delta PI_{t-i}^* + \sum_{j=0}^{r} \psi_j \, \Delta PI_{t-j} + u_t$

Regression #		1	2	3
r, s		0,0	8,8	8,8
Estimation		OLS with HC S.E.	OLS with HC S.E.	GIVE with HC S.E.
$\hat{\beta}_0$		0.00279** (0.00126)	-0.00313 (0.00070)	-0.00262*** (0.00142)
$\hat{\beta}_1$		0.6526* (0.0521)	0.9678* (0.0265)	0.9837* (0.0682)
R^2		0.6661	0.9582	0.7765
$\bar{R}^2$		0.6617	0.9445	0.7032
$\hat{\sigma}$		0.00780	0.00316	0.00730
$\chi^2_{sc}(2)$		25.57*	58.43*	3.21
$\chi^2_{ff}(df)$		0.87	9.69*	12.00*
df		3	3	1
$\chi^2_n(2)$		4.33	2.21	7.71*
$\chi^2_h(1)$		0.53	0.07	10.56*
$\chi^2_{ARCH}(4)$		12.37*	14.55*	0.58
t_{ρ}	*l=3*	-3.12***	-2.85	-4.93*
	l=11	-1.56	-2.36	-2.95
$Z(\hat{\rho}_c)$	*l=3*	-36.87*	-19.59	-78.90*
	l=11	-44.07*	-19.79	-75.42*
$Z(t_{\rho})$	*l=3*	-5.09*	-3.32	-8.98*
	l=11	-5.39*	-3.34	-9.00*

[1] Rejection level of H_0: * at 1%; ** at 5%; *** at 10%.
Source for the critical values of the cointegration tests: Fuller (1976) and Phillips and Ouliaris (1990).
The HSCC estimates of the S.E. were computed using the Newey-West estimator with the Parzen window.
The specification tests are a LM test for residual serial correlation, χ^2_{sc}; a RESET test for functional form χ^2_{ff}; the Jarque-Bera test of normality of the residuals, χ^2_n; a heteroskedasticity test based on the squared of the fitted value, χ^2_h; and a test for autoregressive conditional heteroskedasticity, χ^2_{ARCH}.

cointegration between the two series. Notice that since the LM test for serial correlation fails to detect a problem at order 2 in the IV residuals in strong contrast to the OLS residuals, a shorter lag should suffice to construct the augmented Dickey-Fuller statistic. On the basis of the GIVE estimation of equation (4.43), the intercept is significant only at the 10% level, while the point estimate of β_1 is equal to 0.9837. This estimate is not statistically different from one by the usual Student's distribution even at the 26% significance level. However, this test can be viewed only as suggestive since this coefficient is associated with an integrated regressor so that the estimator of β_1 does not likely follow the asymptotically normal distribution.

Although this was not pursued further in Paquet (1992), the rational expectation hypothesis implies that future expexted inflation will have a direct impact on actual realized inflation, so that it is likely that actual inflation is not weakly exogenous with respect to β_1. In this case, the estimate of β_1 obtained from the estimation of a dynamic specification such as (4.43) would still be biased and inefficient. To deal with this eventuality, Table 4.6 reports estimation results both of a Saikkonen (1991) specification with HSCC estimates of the standard-errors, and of a Phillips and Loretan (1991) specification. We recall that these specifications correct for both serial correlation in the residuals and endogeneity, and that they

Table 4.6- Cointegration tests between PI_t and PI_t^e from efficient modified estimators. (Sample: 1951H2 - 19990H1).[1]

Saikkonen's equation: $PI_t^e = \beta_0 + \beta_1 PI_t + \sum_{i=0}^{8} \psi_{-i} \Delta PI_{t-i} + \sum_{i=1}^{8} \psi_{+i} \Delta PI_{t+i} + v_t$

Phillips' & Loretan's equation: $PI_t^e = \beta_0 + \beta_1 PI_t + \sum_{i=0}^{8} \psi_{-i} \Delta PI_{t-i} + \sum_{i=1}^{8} \psi_{+i} \Delta PI_{t+i} + \sum_{i=1}^{8} \alpha_i (PI_{t-i}^e - \beta_1 PI_{t-i}) + v_t$

Regression #		Saikkonen	Phillips & Loretan I	Phillips & Loretan II
Estimation		OLS with HSCC S.E.	NLS	NLS with HSCC S.E.
$\hat{\beta}_0$		-0.00450** (0.00179)	-0.00112 (0.00085)	-0.00112 (0.00088)
$\hat{\beta}_1$		0.9827* (0.0618)	1.0204* (0.0940)	1.0204* (0.0599)
R^2		0.9369	0.9582	
$\bar{R}^2$		0.9146	0.9761	
$\hat{\sigma}$		0.00411	0.00275	
$\chi^2_{sc}(2)$		32.47*	9.33*	
$\chi^2_{ff}(df)$		5.63**	5.72**	
df		1	1	
$\chi^2_{n}(2)$		0.73	2.98	
$\chi^2_{h}(1)$		1.86	6.43**	
t_{ρ}	$l=11$	-3.60***	-8.62*	
$Z(\hat{\rho}_c)$	$l=11$	-17.18***	-58.21*	
$Z(t_{\rho})$	$l=11$	-3.25***	-8.96*	

[1] Rejection level of H_0: * at 1%; ** at 5%; *** at 10%.
Source for the critical values of the cointegration tests: Davidson and MacKinnon (1993).
The HSCC estimates of the S.E. were computed using the Newey-West estimator with the Parzen window.
The specification tests are a LM test for residual serial correlation, χ^2_{sc}; a RESET test for functional form χ^2_{ff}; the Jarque-Bera test of normality of the residuals, χ^2_{n}; a heteroskedasticity test based on the squared of the fitted value, χ^2_{h}; and a test for autoregressive conditional heteroskedasticity, χ^2_{ARCH}.

are asymptotically equivalent to a ML estimation of the two variables ECM system. Furthermore, we can also conduct tests of interest on the parameters using the standard distributions.

Estimations of both specifications yield stationary residuals, and values of the cointegrating vector consistent with the REH. The estimates of β_1 are not statistically different from one in any case, at the 73% significance level. Notice that because there was evidence of serial correlation and heteroskedasticity left in the residuals for Phillips' and Loretan's specificaton, the NLS covariance matrix was corrected to obtain an HSCC covariance matrix. For both specifications, normality of the error terms could not be rejected at conventional signigicance levels. However, the functional form appeared to be rejected at the 5% significance level. While this could be worrisome if one meant to explain the determination of expected inflations, in this instance, our aim was only to get consistent estimates of the cointegrating coefficients even if the realized inflation series is endogenous. Results from the RESET test are not surprising since the modified specifications were certainly not intended to provide an explanatory equation of determination of the regressand.

4.4 Multivariate Time Series Processes

4.4.1 Introduction

We have discussed so far many relevant econometric issues concerning individual dynamic equations, such as the univariate representation of a single time series and the empirical modelling of a dynamic relationship between several variables. The latter required that all but one of them be weakly exogenous regarding the parameters of interest. There are some instances for which one would like to consider jointly the dynamic determination of several time series. One possibility would require the specification of a structural dynamic simultaneous equation model. In this case, one needs to classify the system's variables as endogenous or exogenous, and to specify the restrictions on the parameters or on the covariances to guarantee the identification of the model's parameters. These steps must precede estimation and inference.

Sims (1980) has proposed a different approach generalizing the *AR* univariate method of time series to a multivariate context. Vector autoregressions (*VAR*) essentially consist in modelling jointly each variable of the system as a linear function of lagged values of all the variables.[34] While *VAR*'s were originally used as an atheoretical approach to characterize macroeconomic data, recent developments by Sims (1986) and others led to structural vector autoregressions (*SVAR*). This section discusses in turn how *VAR* and *SVAR* are applicable with stationary and with integrated variables. It considers also how a reparameterization of a *VAR* into a vector error-correction model (*VECM*) can be used to test the number of cointegrating relationships between integrated regressors, to estimate cointegrating vectors, and to conduct inference.

4.4.2 Vector autoregressive models (VAR)

4.4.2.1 The case of stationary regressors.

4.4.2.1.1 Set-up and estimation of a VAR

Let us consider a column vector of N stationary variables $\mathbf{X}_t = [x_{1t}, \ldots, x_{Nt}]'$, the vector autoregressive system $VAR(p)$ for the t^{th} observation is defined as:

---

[34] While we can also model vector autoregressive and moving average (*VARMA*) systems, these have not caught on as much in empirical work. Readers who are interested in *VARMA* representation should consult Granger and Newbold (1986), or Harvey (1981, 1990).

$$\mathbf{X}_t = \boldsymbol{\delta} + \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{X}_{t-2} + \cdots + \boldsymbol{\Phi}_p \mathbf{X}_{t-p} + \boldsymbol{\varepsilon}_t = \boldsymbol{\Phi}(L)\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{4.44}$$

$$\text{where } \boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_N \end{bmatrix}_{N\times 1}, \quad \boldsymbol{\Phi}_i = \begin{bmatrix} \phi_{11,i} & \cdots & \phi_{1N,i} \\ \vdots & \ddots & \vdots \\ \phi_{N1,i} & \cdots & \phi_{NN,i} \end{bmatrix}_{N\times N} \quad \text{for } i = 1,\ldots,p \quad ,$$

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{N\times 1}, \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{bmatrix}_{N\times N} \right) \equiv (0, \Sigma_\varepsilon) \ ,$$

and $\boldsymbol{\varepsilon}_t$ is assumed to be a vector white noise process, i.e. $E(\boldsymbol{\varepsilon}_t\, \boldsymbol{\varepsilon}_s') = 0$, for any $t \neq s$.

Equation (4.44) is nothing but a system of N dynamic linear equations. In particular, the n^{th} row of the equations system (4.44) can be written as

$$x_{nt} = \delta_n + \sum_{i=1}^{p} \sum_{j=1}^{N} \phi_{nj,i}\, x_{jt-i} + \varepsilon_{nt}, \tag{4.45}$$

that is a linear regression equation of the n^{th} variable on a constant and p lags of all the N variables of the system. A typical coefficient $\phi_{nj,i}$ denotes the effect of the i^{th} lag of the j^{th} variable on the n^{th} variable. Since a VAR(p) system is an extension of the univariate $AR(p)$ process, its stationarity requires that both the multivariate mean and covariances be bounded, and that the autocovariance matrices between $\mathbf{X}_t$ and $\mathbf{X}_{t+\tau}$ be only functions of the time interval τ, but not of time itself. A necessary and sufficient condition for a $VAR(p)$ to be stationary is that the characteristic equation of the system have all its roots with modulus greater than one. The characteristic roots are the solutions of the following equation:

$$\left| \mathbf{I}_{N\times N} - \boldsymbol{\Phi}_1 z - \boldsymbol{\Phi}_2 z^2 - \cdots - \boldsymbol{\Phi}_p z^p \right| = 0 . \tag{4.46}$$

The estimation of a $VAR(p)$ model for N variables involves $N + p\,N^2$ coefficients and $N(N+1)/2$ error variances and covariances. Given a lag-length p, because a VAR system allows the innovations of each equation to be contemporaneously correlated, this is analogue to a seemingly unrelated regressions model. But, since all the system's equations share the same matrix of regressors, the seemingly unrelated regressions estimation (SURE) amounts to the application of OLS separately on each equation. Zellner (1962) also proved the consistency and the efficiency of OLS in such systems. Once the coefficients have been estimated, the OLS residuals can be used to construct the Gaussian maximum likelihood estimate of the error covariance matrix $\hat{\Sigma}_\varepsilon$, whose elements are:

$$\hat{\sigma}_{nl} = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{nt}\hat{\varepsilon}_{lt} \qquad \forall n,l = 1,\ldots,N. \qquad (4.47)$$

We, therefore, have:

$$\hat{\Sigma}_{\varepsilon} = T^{-1}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t'. \qquad (4.48)$$

4.4.2.1.2 Choice of a VAR's lag-length and diagnostic checking

The set-up and estimation of a *VAR* system require the determination of its order or lag-length p. Even though tools such as cross-correlograms that picture the cross-correlations between different variables at different lags can help identifying the proper lag, in practice they are seldomly used for this purpose. Instead, researchers use hypothesis testing on the significance of additional lags, and data-dependent information criteria. When applying either of these tools, care must be taken however so that a common sample size is employed to compare the adequacy of different lag length. In our exposition below, we suppose that there are in effect T observations available for estimation of each alternative specification.

If we wish to test the null hypothesis that the longest lag of the system is p against the alternative that it is $p+s$, we would follow the steps below.

1. Estimate the *VAR*(p) and get the estimate of the covariance matrix of its errors, i.e. $\hat{\Sigma}_{\varepsilon}(p)$.

2. Estimate the *VAR*($p+s$) and get the estimate of the covariance matrix of its errors, i.e. $\hat{\Sigma}_{\varepsilon}(p+s)$.

3. To test $H_0 : \Phi_{p+1} = \Phi_{p+2} = \cdots = \Phi_{p+s} = 0$, one constructs the following likelihood ratio (*LR*) statistic:

   $$T\left(\log\left|\hat{\Sigma}_{\varepsilon}(p)\right| - \log\left|\hat{\Sigma}_{\varepsilon}(p+s)\right|\right) \overset{a}{\sim} \chi^2(sN^2). \qquad (4.49)$$

   Notice that this distribution is invoked on asymptotic ground. Because of the potentially large number of parameters under both the null hypothesis and the alternative hypothesis, this test is commonly performed with $s = 1$.

Another way to select the lag length consists in choosing the value of p that minimizes either of the following criteria:

-the Akaike information criterion: $AIC(p_j) = \log\left(\left|\hat{\Sigma}_{\varepsilon}(p_j)\right|\right) + \dfrac{2N^2p_j}{T}, \qquad (4.50)$

-Schwarz information criterion: $$SC(p_j) = \log\left(\left|\hat{\Sigma}_\varepsilon(p_j)\right|\right) + \frac{N^2 p_j \log(T)}{T}, \qquad (4.51)$$

-Hannan and Quinn criterion:

$$HQ(p_j) = \log\left(\left|\hat{\Sigma}_\varepsilon(p_j)\right|\right) + \frac{2N^2 p_j \log(\log(T))}{T}, \qquad (4.52)$$

over the j alternatives being considered, i.e. $p_j \in \{0,1,\ldots,p_{\max}\}$. These criteria are indices expressing a trade off between the goodness-of-fit and the parsimony of the specification. The former is captured by the first term of the expressions, while the latter is accounted by the second term that penalizes for too many parameters relative to the sample size.[35] What distinguishes the three criteria above are the penalty term involved in their reckoning.

Typically, for a given p_j, the values of these criteria will be ranked as $AIC(p_j) \leq HQ(p_j) \leq SC(p_j)$. That is, the Schwarz criterion penalizes the most the inclusion of extra lags, while Akaike's has the lowest penalty. For these reasons, all criteria will not necessarily suggest the same lag length. In fact, practical experience shows that the Schwartz criterion will often choose too small an order for the *VAR* system.

Once a value of p is chosen from some relevant criterion, it remains crucial to apply a diagnostic check procedure about the statistical adequacy of the maintained specification. In particular, too small a p will lead to inconsistent and biased estimates of the *VAR* coefficients. Missing dynamics would likely lead to time dependence in the residuals, which would violate the assumption of ε_t being a vector white noise process. One should examine the behaviour of the sample ordinary and partial autocorrelations of each equation's residual. The application of *LM* tests for serial correlation should also be applied on each residual.

4.4.2.1.3 Applied tools specific to VAR models.

Because they do not rely upon *a priori* identifying restrictions, atheoretical *VAR* models had been proposed as a potentially preferable method to large dynamic simultaneous equations system. As such, they could remain essentially a multivariate atheoretical tool to characterize macroeconomic data. In particular, *VAR* can be used to construct optimal conditional forecasts. One problem is that they often involve a large number of parameters that may not be estimated very precisely. Doan, Litterman and Sims (1984), and Litterman (1980, 1986) have proposed to incorporate some prior distributions of the coefficients following a Bayesian methodology that uses the data to update this distribution. Applications of Bayesian vector autoregression (*BVAR*) have often lead to better out-of-sample forecasts than regular *VAR*. Further discussion on *BVAR* is however beyond the scope of this guide.[36]

---

[35] In fact, the first term of the information criteria amounts to a *LR*-test of a $VAR(p_j)$ against a $VAR(0)$, with the critical values being determined by the penalty term, instead of some value based on a χ^2-distribution.

Besides forecasting, *VAR* systems are also used for innovation accounting exercises. These analyze the reaction of the system's variables following a disturbance of say one standard-deviation on each innovation. Impulse response functions (*IRF*) of each variable are plotted to follow the impact of a shock, say, on the j^{th} variable after 1 period, 2 periods, and so forth. Impulse response functions can be computed from rewriting the stationary *VAR*(p) system in its non-explosive infinite vector moving average (*VMA*) representation. For instance, the *VMA*(∞) corresponding to the *VAR*(p) of equation (4.44) is given by:

$$X_t = \Psi(1)\delta + \sum_{i=0}^{\infty} \Psi(L)\,\varepsilon_t = \mu + (I + \Psi_1 L + \Psi_2 L^2 + \cdots + \Psi_p L^p)\,\varepsilon_t \tag{4.53}$$

where $\Psi(L) = (I - \Phi_1 L - \cdots - \Phi_p L^p)^{-1}$ and $\mu = E(X_t) = \Psi(1)\delta$. This representation expresses all variables as functions of the current and lagged innovations of the system. In particular, $\partial X_{t+i} / \partial \varepsilon_t' = \Psi_i$ is an $N \times N$ matrix whose $(k, j)^{th}$ element measures the response of the k^{th} variable at date $t+i$ to a shock on the j^{th} variable that would occur at date t. The impulse response function is defined as the values of these $(k, j)^{th}$ elements as a function of i.

Estimates of the Ψ_i matrices involve highly non-linear functions of the estimates of the $\phi_{n,j,i}$ coefficients, which makes unfeasible direct linear computations of the standard-errors of the impulse responses. Instead, following Runkle (1987), Monte Carlo simulations can be used very effectively to reckon confidence intervals of the impulse response functions.[37, 38]

---

[36] Racette and Raynauld (1992) is an interesting application of the BVAR method to Canadian data. The comments by Boothe (1992) offer an entertaining and critical assessment of the BVAR methodology.

[37] Runkle (1987) showed that these confidence intervals can be computed using either a Normal approximation, or a technique called bootstrapped resampling. For instance, in the latter case, one can randomly sample with replacement the actual residuals to draw 500 replications or series of innovations. Then, these are used as shocks to the *VAR* to generate 500 artificial series for the *N* variables. After reestimating the VAR for each set of artificial variables, he or she computes the corresponding impulse response functions. Finally, standard errors of the coefficients making up the estimated Ψ_i matrices can be computed from the 500 generated estimates. This is fairly easy to do using the RATS software.

[38] Mittnick and Zadrozny (1993) have also derived formulae to compute asymptotic covariance matrices of impulse responses and variance decompositions.

Typically, because of the large number of parameters estimated in a *VAR*, confidence intervals at conventional statistical levels (e.g. 95%) tend to be fairly large.

One complication that arises in practice is that the innovations are generally contemporaneously correlated. Since the nature of a *VAR* system implies one period delay before a shock on one variable is propagated onto other variables, if the errors of the x_j's are contemporaneously correlated, it is not possible without further information to disentangle the effect specific to each innovation. To deal with this problem, one must rely on some transformation of the *VAR* in a way that make the transformed errors orthogonal (or uncorrelated). One such method involves using a Choleski decomposition to find some lower triangular matrix P to pre-multiply the original *VAR*. Then, equation (4.44) would become:

$$\mathbf{P}\,\mathbf{X}_t = \mathbf{P}\,\delta + \Phi_1\mathbf{P}\,\mathbf{X}_{t-1} + \Phi_2\,\mathbf{P}\,\mathbf{X}_{t-2} + \cdots + \Phi_p\,\mathbf{P}\,\mathbf{X}_{t-p} + \mathbf{P}\,\varepsilon_t, \qquad (4.54)$$

where P is the unique lower triangular matrix that solves $\Sigma_\varepsilon = (\mathbf{P}^{-1})(\mathbf{P}^{-1})'$ such that $E[\mathbf{P}\,\varepsilon_t\,\varepsilon_t'\,\mathbf{P}'] = \mathbf{P}\,E[\varepsilon_t\,\varepsilon_t']\mathbf{P}' = \mathbf{P}\,\Sigma_\varepsilon\,\mathbf{P}' = \mathbf{I}$.[39] This transformation then implies the following *VMA*(∞):

$$\mathbf{X}_t = \mathbf{P}\,\mu + \sum_{i=0}^{\infty} (\Psi_i\mathbf{P}^{-1})(\mathbf{P}\,\varepsilon_{t-i}) = \mu^* + \sum_{i=0}^{\infty} \Psi_i^*\ \varepsilon_{t-i}^*, \qquad (4.55)$$

where $\varepsilon_t^* = \mathbf{P}\,\varepsilon_t$ is the vector of orthogonalised innovations with a variance-covariance matrix given by the identity matrix, i.e. $\Sigma_{\varepsilon^*} = E[\,\varepsilon_t^*\varepsilon_t^{*\prime}] = \mathbf{I}$.

While it is helpful, the orthogonalisation procedure is not unique and is function of the ordering of the variables, i.e. the position of each variable in the $\mathbf{X}_t$ vector. The choice of the ordering may be somewhat arbitrary; or in particular instances, it may be rationalised by some structural interpretation of the *VAR*. For example, if one were interested in the *IRF* of a *VAR* that included both the US and Canadian interest rates, it might seem reasonable to set the first element of $\mathbf{X}_t$ to be the US interest rate. We would in effect be assuming that US rates movements precede Canadian rate innovations, but that the latter have no effects on US rates.

---

[39] As explained in Keating (1992), a Choleski decomposition of the residual's variance-covariance matrix for a *N* variables VAR model ordered from variable *1* to variable *N* is algebraically equivalent to estimating the following *N* equations by OLS. Namely, $\varepsilon_{1t} = \varepsilon_{1t}^*$; $\varepsilon_{2t} = p_{12}\,\varepsilon_{1t} + \varepsilon_{2t}^*$; $\varepsilon_{3t} = p_{31}\,\varepsilon_{1t} + p_{32}\,\varepsilon_{2t} + \varepsilon_{3t}^*$; ...; $\varepsilon_{Nt} = p_{N1}\,\varepsilon_{1t} + p_{N2}\,\varepsilon_{2t} + \ldots + p_{NN-1}\,\varepsilon_{N-1t} + \varepsilon_{Nt}^*$. The p_{ij} are the elements of the lower triangular matrix P; the diagonal elements are 1's. Obviously, changing the ordering of the variables changes the value of matrix P.

Cooley and Leroy (1985) have precisely criticised so called atheoretical *VAR*'s, arguing that if the Choleski decomposition was to be interpreted as non-structural, the ε_t^*'s are not structural innovations and should not be interpreted as such. They are then typically linear combinations of the true structural disturbances. Furthermore, they argue that the Choleski shocks are not really atheoretical. Rather, the Choleski decomposition imposes a particular recursive contemporaneous structure to the underlying model that may or may not make economic sense, and which is function of the ordering of the variables. This is why they also reject the practice, which consists to emphasise conclusions that are so-called robust to different orderings of the variables. The only orderings that should be of interest should be justified as implying some economically sensible structure. Cooley and Leroy's (1985) criticisms of atheoretical *VAR*'s have led to recent developments, which make explicit use of identifying restrictions to give a structural interpretation to the use of *VAR*'s. The next subsection will address the structural interpretation of *VAR* models in more details.

Another tool associated with the analysis of *VAR* systems is the variance decomposition function, which attribute the contribution of the individual shocks to each variable's forecast error variance. Here again, the issue of orthogonalisation applies as it did for computing impulse response functions. Referring to the *VMA* representation (4.55), we can write the optimal forecast of X_{t+s} conditional on information known as of date t, about the N variables of the system. This is,

$$E_t[X_{t+s}] = E[X_{t+s} \mid X_t, X_{t-1}, ...] = \mu^* + \sum_{i=s}^{\infty} \Psi_i^* \ \varepsilon_{t+s-i}^* . \tag{4.56}$$

Hence, the s-step ahead forecast error is defined as:[40]

$$X_{t+s} - E_t[X_{t+s}] = \sum_{i=0}^{s-1} \Psi_i^* \varepsilon_{t+s-i}^* . \tag{4.57}$$

Conformably, the variance-covariance matrix of the s-period ahead forecast error is given by:

$$\begin{aligned} E\left\{(X_{t+s} - E_t[X_{t+s}])(X_{t+s} - E_t[X_{t+s}])'\right\} &= \sum_{i=0}^{s-1} \Psi_i^* E\left(\varepsilon_{t+s-i}^* \varepsilon_{t+s-i}^{*\prime}\right) \Psi_i^{*\prime} \\ &= \sum_{i=0}^{s-1} \Psi_i^* \Psi_i^{*\prime} . \end{aligned} \tag{4.58}$$

---

[40] It is also possible to define a s-step ahead forecast error as $X_t - E_{t-s}[X_t] = \Sigma_i \Psi_i^* \varepsilon_{t-i}^*$, where $i \in [0, s\text{-}1]$.

Letting $\psi^{*}_{mj,i}$ be the $(m, j)^{th}$ element of matrix Ψ^{*}_{i}, the s-step-ahead forecast variance for the m^{th} variable is given by:

$$E(x_{mt} - E_t x_{mt+s})^2 = \sum_{i=0}^{s-1} \sum_{k=1}^{N} \psi^2_{mk,i}. \tag{4.59}$$

The s-step-ahead percentage of forecast variance for the m^{th} variable that is associated to the j^{th} shock thus defines the variance decompositions (*VD*) as a function of s, i.e.

$$VD(m,j,s) = \frac{\sum_{i=0}^{s-1} \psi^2_{mj,i}}{\sum_{i=0}^{s-1} \sum_{k=1}^{N} \psi^2_{mk,i}} \times 100. \tag{4.60}$$

Similarly to the impulse responses, the variance decompositions of the orthogonalised innovations are generally sensitive to the ordering of the variables in the *VAR* system. Standard-errors of *VD*'s can also be computed by Monte Carlo integrations or by Normal approximation.[41]

Finally, another use of the *VAR* system is to test the null hypothesis of Granger non-causality of a variable $x_{j,t}$ on another variable, say x_{lt} in a bivariate (N=2) model.[42] In the context of a *VAR*, this would mean that movements in any lagged value of x_{jt} would provide no information about current movements of x_{lt}. Notice that the concept of Granger non-causality bears no implication about some logical causal link between two variables, it is nothing more than a statement about statistical precedence. The null hypothesis of Granger non-causality of the variable x_{lt} on x_{jt} can be set as:

$$H_0 : \text{Granger non-causality of } x_{jt} \text{ on } x_{lt} : \phi_{lj,1} = \phi_{lj,2} = \cdots = \phi_{lj,p} = 0.$$

The simple way to test this joint null hypothesis in a *VAR*(p) context is to construct an F-test that would asymptotically follow an $F(p, T\text{-}(1\text{+}pN))$ distribution.[43]

Granger non-causality is distinct from predeterminedness and many concepts of exogeneity.[44] It should not serve much more than as an atheoretical characteristic of the data.

---

[41] See Runkle (1987), Mittnick and Zadrozny (1993), or Hamilton (1994).

[42] The concept of Granger non-causality can also be extented to a multivariate ($N > 2$) *VAR* system for testing whether a subset of variables is Granger causing another subset of variables, provided that the two subsets exhaust all variables in the system and that they are mutually exclusive. See Hamilton (1994).

[43] Geweke (1984) surveys and discusses alternative test statistics of non-causality.

[44] A variable is said to be predetermined if it is independent of contemporaneous and future errors in an equation. Granger non-causality is a necessary condition for strong exogeneity, but it is not sufficient. See Engle *et al.* (1983) and

Also, results from such tests will be sensitive to the omission of relevant variables from the *VAR* system and the power of such tests may be sensitive to the lag-length order *p*.

4.4.2.1.4 Issues in structural *VAR's*

In response to critics about atheoretical *VAR*'s, recent papers have seeked to combine structural and multivariate time-series models.[45] Consider the following structural dynamic simultaneous equations system:

$$\mathbf{A}\,\mathbf{X}_t = \mathbf{B}(L)\,\mathbf{X}_{t-1} + \mathbf{u}_t, \tag{4.61}$$

where $\mathbf{A}$ is an $N{\times}N$ matrix representing the contemporaneous ties between the variables of the system, $\mathbf{B}(L)$ is a p-order matrix polynomial in L, and $\mathbf{u}_t$ is an $N{\times}1$ vector of serially uncorrelated structural innovations with a variance matrix given by $E(\mathbf{u}_t\mathbf{u}_t') = \Sigma_u$. The structural system (4.61) implies moreover the following reduced form:

$$\mathbf{X}_t = \mathbf{A}^{-1}\,\mathbf{B}(L)\mathbf{X}_{t-1} + \mathbf{A}^{-1}\,\mathbf{u}_t = \Phi(L)\,\mathbf{X}_{t-1} + \varepsilon_t\,, \tag{4.62}$$

with $E(\varepsilon_t\varepsilon_t') = \Sigma_\varepsilon$, and expresses all the system's variables in terms of its own lags and the lags of the other variables of the system. Therefore, equation (4.62) is in effect a *VAR*(p) system with restrictions implied by the underlying structural model.

From the estimation of the *VAR* system (4.62) as an unrestricted model, we can write its corresponding *VMA* representation as:

$$\mathbf{X}_t = [\mathbf{I} - \Phi(L)L]^{-1}\,\varepsilon_t = [\mathbf{I} - \mathbf{A}^{-1}\mathbf{B}(L)L]^{-1}\,\varepsilon_t = \Psi(L)\,\varepsilon_t\,. \tag{4.63}$$

But, if we are concerned with the dynamic responses of a variable to past structural innovations on a given variable, we would instead be interested in the *VMA* representation of equation (4.61), namely:

$$\mathbf{X}_t = [\mathbf{A} - \mathbf{B}(L)L]^{-1}\,\mathbf{u}_t = \Psi^*(L)\,\mathbf{u}_t\,. \tag{4.64}$$

Hence, these two *VMA* would typically imply very different dynamics since the errors of the unrestricted *VAR*, i.e. ε_t, are different from the unobserved fundamental disturbances of the structural system, i.e. $\mathbf{u}_t$. However, if matrix $\mathbf{A}$ could somehow be identified, we could derive

---

Cooley and LeRoy (1985).

[45] For examples, see: Sims (1986), Bernanke (1986), Blanchard and Watson (1986), Blanchard and Quah (1989), Keating (1990, 1992), King, Plosser, Stock and Watson (1991), and Christiano and Eichenbaum (1992, 1994).

the structural innovations from the *VAR*'s residuals since $\varepsilon_t = \mathbf{A}^{-1}\,\mathbf{u}_t$. This expression could be substituted into equation (4.63), which would thus reveal the dynamics of the structural innovations.[46] An assessment of the model's adequacy could then be performed by checking whether the dynamic patterns revealed by both the impulse responses and the variance decompositions are consistent with the underlying model used for identification.

A first method for identifying the underlying structural model consists in specifying theoretical restrictions on the contemporaneous structure of the model. Indeed, the OLS estimation of the unrestricted *VAR* in equation (4.62) yields estimates for the variance-covariance matrix of the residuals, i.e. $\Sigma_\varepsilon = E[\varepsilon_t\, \varepsilon_t']$, and since we know how the *VAR*'s errors are related to the structural innovations, the following equation must hold:

$$\Sigma_\varepsilon = E\left[\varepsilon_t\,\varepsilon_t'\right] = E\left[\mathbf{A}^{-1}\,\mathbf{u}_t\,\mathbf{u}_t'\,\mathbf{A}^{-1\prime}\right] = \mathbf{A}^{-1}\,\Sigma_u\,\mathbf{A}^{-1\prime}\,. \tag{4.65}$$

There are N^2 unknown elements in $\mathbf{A}$, $N(N+1)/2$ unknown unique elements in Σ_u, and $N(N+1)/2$ unique and estimated elements for Σ_ε. In order to identify the structural parameters of the system, we therefore require at least N^2 identifying restrictions on $\mathbf{A}$ and Σ_u.[47] Typically, Σ_u is defined as being a diagonal matrix, with variances of each element being normalised to unity, so that $\Sigma_u = \mathbf{I}$. At this point, we are still left with $(N^2 - N(N+1)/2) = N(N+1)/2$ missing identifying restrictions.

One possibility is to consider a recursive or Wold chain causal contemporaneous structure between the system's variables. Namely, for a given ordering of the variables in the system, suppose one assumes the following: the first variable responds only to its own innovations; the second variable responds contemporaneously to innovations in the first variable as well as to its own structural shocks; the third variable contemporaneously responds to its idiosyncratic disturbances, and to shocks on the first and second variables; and so forth. In effect, this structure would impose $N(N+1)/2$ identifying restrictions by normalising the N diagonal elements of $\mathbf{A}$ to be *1*'s, and by imposing that the $N(N-1)/2$ upper diagonal elements of matrix $\mathbf{A}$ be equal to zero, so that $\mathbf{A}$ is a lower triangular matrix. This identifying structure amounts therefore to the Choleski decomposition of the *VAR*'s errors covariance matrix since $\Sigma_\varepsilon = (A^{-1})(A^{-1})'$. This particular structure has been used recently by Christiano and Eichenbaum (1992, 1994) to identify monetary policy shocks.

---

[46] Indeed, $[\mathbf{A} - \mathbf{B}(L)\,L]^{-1} = [\mathbf{I} - \mathbf{A}^{-1}\,\mathbf{B}(L)\,L]^{-1}\,\mathbf{A}^{-1}$.

[47] If we had instead considered a generalisation of equation (4.62), such as $\mathbf{A}\,\mathbf{X}_t = \mathbf{B}(L)\,\mathbf{X}_{t-1} + \mathbf{C}\,\varepsilon_t$, with $\mathbf{C}$ having N^2 elements, we would then have required $2\,N^2$ identifying restrictions on $\mathbf{A}$, $\mathbf{C}$, and Σ_u. See Bernanke (1986), and Giannini (1992).

Other researchers (e.g. Bernake (1986), Blanchard and Watson (1986), Sims (1986), Keating (1992)) have imposed $N(N+1)/2$ contemporaneous identifying restrictions on matrix A, without requiring it to be lower triangular. In a particular *3* variables *VAR*, one might, for instance, argue that economic theory suggests a contemporaneous relationship like:

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} = \mathbf{A}^{-1}\mathbf{u}_t = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{21} & 1 & \alpha_{23} \\ 0 & \alpha_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}. \tag{4.66}$$

In this case, the identification device does not correspond the Choleski decomposition. Estimates for matrix A then need to be recovered by solving the nonlinear system of equations implied by equation (4.65).

Another way to achieve identification of a *SVAR* is to impose long-run restrictions in place of, or in combination with contemporaneous identifying restrictions. This avenue has been pursued by Shapiro and Watson (1988), Blanchard and Quah (1989), King, *et al.* (1991), and Keating (1992). For instance, economic theory might suggest particular values for the long-run multipliers of the structural system (4.61). E.g., in a model with monetary and real variables, one might want to impose that cumulative effects of a monetary shock on a real variable be zero in the long run.

To discuss the role of long-run identifying restrictions, let $X_t = X_{t-1} = X^*$ in the long-run, so that by setting L equal to 1 in equation (4.61), we obtain:

$$[\mathbf{A} - \mathbf{B}(1)]\,\mathbf{X}^* = \mathbf{u}, \tag{4.67}$$

where $\mathbf{B}(1) = \Sigma_{i=1}^{p} \mathbf{B}_i$, and its long-run *VMA* representation is given by:

$$\mathbf{X}^* = [\mathbf{A} - \mathbf{B}(1)]^{-1}\mathbf{u} = \Psi^*(1)\,\mathbf{u}. \tag{4.68}$$

Equation (4.68) implies that $\mathbf{A}^{-1} = [\mathbf{I} - \mathbf{A}^{-1}\mathbf{B}(1)]\,\Psi^*(1)$, where $\mathbf{A}^{-1}\,\mathbf{B}(1) = \Phi(1)$ is the total sum of the *VAR* matrices of coefficients. Substituting out $\mathbf{A}^{-1}$ in equation (4.65) then yields:

$$[\mathbf{I} - \Phi(1)]^{-1}\,\Sigma_\varepsilon\,[\mathbf{I} - \Phi(1)]^{-1} = \Psi^*(1)\,\Sigma_u\,\Psi^*(1)'. \tag{4.69}$$

By a reasoning similar to that used with contemporaneous identifying restrictions, with $\Sigma_u = I$, and the elements of the main diagonal of $\Psi^o(1)$ being set to one, equations (4.69) implies that $N(N-1)/2$ additional identifying restrictions are still needed. These can be seen as restrictions on the sum of impulse responses. One possibility is to impose a long-run Wold causal chain by making $\Psi^o(1)$ lower triangular. Another is to impose the additional restrictions without requiring the recursivity specific to a particular ordering.

There is finally another avenue for identifying *SVAR* models that aims at uncovering deep structural parameters under the rational expectations hypothesis. Keating (1990) has proposed to rely on rational expectations restrictions that a particular structure implies for the reduced form (or *VAR* representation) rather than using contemporaneous exclusion restrictions. He shows in fact that the latter approach might even lead to inconsistent estimates in many circumstances. Interested readers are referred to the published paper for further details.

We have discussed identification of a *SVAR* by referring essentially to the so called order conditions for identification. However, a rank condition is also required for identification. For instance, if the structure of matrix A was such that it did not preclude the possibility that some of its rows be linearly dependent, we would still be left with an unidentified model.[48] Giannini (1992) and Hamilton (1994) discuss the procedure to verify the validity of the rank condition in a *SVAR*.[49] Note that in the case where the Wold causal chain served as identifying restrictions (i.e. when using the Choleski decomposition), both the rank and order conditions are satisfied.

Our discussion has also focused on just-identified *SVAR* for which the number of identified parameters is exactly equal to the number of unique elements in the variance-covariance matrix of the *VAR* residuals, Σ_ε. In this case, the OLS estimation of a correctly specified *VAR* is efficient. However, there are cases where there would be fewer parameters than the number of unique elements in Σ_ε. With overidentified *SVAR*, other estimation procedures such as FIML or GMM are preferable and are discussed in Watson (1993).

---

[48] For example, even though the order condition is satisfied, the rank condition is violated in the following 4 variables *VAR* for which the first, second, third, and fourth row of matrix A are given respectively by: $(1, \alpha_{12}, 0, \alpha_{14})$, $(\alpha_{21}, 1, 0, \alpha_{24})$, $(0, \alpha_{32}, 1, 0)$, and $(0, 0, \alpha_{43}, 1)$.

[49] Giannini (1992) gives also RATS codes to apply such a procedure.

Another important question is how sensitive the tools used in conjunction with *VAR* and *SVAR* models are to potential misspecifications. Keating (1990) mentions that existing *SVAR*'s assume time-invariant parameters. He also underlines how the omission of important variables may adversely affect the results. Braun's and Mittnik's (1993) quantitative illustrations suggest omitted variables and underfitting of the lag-length can be quite damaging for consistency of the estimates.[50] The consequences of these misspecifications are somewhat reduced for the impulse responses and the prediction-errors variance estimates when taking parameter uncertainty into account, i.e. when constructing confidence intervals around the point estimates. This is not so however for the variance decompositions, which are then ratios of inconsistent estimates. Hence, as other macroeconometric techniques discussed in this guide, the econometrics of *SVAR* models cannot abstract from issues of specification adequacy.

4.4.2.2 Generalisation to the case of I(1) variables.

We have so far discussed the econometrics of *VAR* and *SVAR* systems that involved only stationary variables. Some generalisations and some specificities apply when a *VAR* includes integrated time series. In fact, provided that each equation of the *VAR* system is balanced so that the equations' innovations are $I(0)$, one can estimate the coefficients of the system in the usual manner. Namely, the equations of the *VAR* are *ADL* models that we studied earlier in this chapter. Inference is however no so unequivocal.

For convenience, let us rewrite the N variables $VAR(p)$ model given by equation (4.44):

$$\mathbf{X}_t = \delta + \Phi_1 \mathbf{X}_{t-1} + \Phi_2 \mathbf{X}_{t-2} + \cdots + \Phi_p \mathbf{X}_{t-p} + \varepsilon_t \,. \tag{4.70}$$

Then, adding and subtracting $\left[\Sigma_{j=1}^{p-1}\left(\Sigma_{i=j}^{p}\ \Phi_{i+1}\right)\mathbf{X}_{t-j}\right]$ on the RHS of equation (4.70), we obtain the following reparameterisation:

$$\mathbf{X}_t = \delta + \Lambda_1 \Delta\mathbf{X}_{t-1} + \Lambda_2 \Delta\mathbf{X}_{t-2} + \cdots + \Lambda_{p-1} \Delta\mathbf{X}_{t-p+1} + \Xi\, \mathbf{X}_{t-1} + \varepsilon_t, \tag{4.71}$$

where the $N \times N$ matrices of coefficients are:

$$\Lambda_i = -(\Phi_{i+1} + \Phi_{i+2} + \cdots + \Phi_p) \qquad \text{for } i = 1, \ldots, p-1, \tag{4.72a}$$

---

[50] They also consider the consequences of ignoring *MA* behaviour of the *VAR* residuals. Depending on the sample size available, this type of misspecification can be addressed by adjusting the order of the *VAR* under consideration.

and

$$\Xi = \Phi_1 + \Phi_2 + \cdots + \Phi_p = \Phi(1). \tag{4.72b}$$

It follows from equation (4.46) that the *VAR* in levels given by equation (4.71) contains at least one unit root if the following determinant is zero:

$$\left| \mathbf{I}_{N \times N} - \Phi_1 - \Phi_2 - \cdots - \Phi_p \right| = 0 . \tag{4.73}$$

A first possibility is that this is so because $\Xi = \Phi(1) = \mathbf{I}_{N \times N}$. In this case, equation (4.72) amounts to a *VAR*(*p-1*) model for the first-differences of $\Delta \mathbf{X}_t$:

$$\Delta \mathbf{X}_t = \delta + \Lambda_1 \Delta \mathbf{X}_{t-1} + \Lambda_2 \Delta \mathbf{X}_{t-2} + \cdots + \Lambda_{p-1} \Delta \mathbf{X}_{t-p+1} + \varepsilon_t . \tag{4.74}$$

The reason why the *VAR* in levels given by equation (4.70) is equivalent and non-spurious is that the inclusions of lagged values of both the dependent variables and the regressors in each equation is equivalent to first-differencing all the system's variables. The unit roots of the system are in effect implicitly estimated. Provided that the lag-length is appropriate, estimation of the *VAR* in levels or in first-differences is straightforward and can be carried out by OLS.

A second possibility is that equation (4.73) holds, which indicates the presence of one or more unit roots in the *VAR* in levels, but that there are cointegrating relationships between some the model's *I*(*1*) variables. In this case, the OLS estimation of the *VAR* in levels remains an option, while that of equation (4.74) would be biased and inconsistent since it does not explicitly or implicitly takes account of the existing cointegrating relationships that exist.[51]

Even though the OLS estimation of a *VAR* with *I*(*1*) regressors in levels is still applicable as long as the errors are *I*(0), complications arise at the inference stage. That is because the least-squares estimates of integrated regressors' coefficients do not necessarily follow a standard asymptotic distribution. To perform test statistics about null hypothesis involving coefficients of *I*(*1*) regressors, while referring to usual tables of critical values, one needs to investigate whether any of the rules presented in section 4.3.3.3 apply. As shown by Sims, Stock and Watson (1990), inference about parameters of *I*(*1*) variables can be conducted with standard asymptotic distribution if there exists one or many reparameterizations of the equation that express the restrictions to be tested on stationary regressors. This is true even for some cross equation restrictions. In particular, Sims, Stock and Watson's (1990) result implies that the following hypothesis tests on levels *VAR* systems can be carried with the conventional distributions:

---

[51] The next section will discuss a procedure that explicitly estimates cointegrating relationships in a vectorial error-correction framework.

1. Tests of lag length restrictions on a particular x_{nt};
2. Tests on a subset of the lagged levels of a particular x_{nt}, including individual coefficient tests of significance.

Tests of Granger non-causality of an $I(1)$ variable x_{jt} on another $I(1)$ variable x_{nt} generally follow nonstandard distributions. In particular, Toda and Phillips (1993a) and Ohanian (1988) have shown that the use of a conventional χ^2 distribution will lead too often to the rejection of Granger non-causality of an independent random walk. Toda and Phillips (1993b) showed that:

1. If there is no cointegration amongst the $I(1)$ variables of a model, Granger non-causality tests in levels *VAR*'s have an asymptotic unit root distribution that is nuisance parameters free. Provided that one computes the appropriate critical values, the test is applicable on the levels *VAR*. In this case, it is also possible to conduct the non-causality test on the differences *VAR*'s, using conventional distributions. The latter might even be more powerful in finite samples as it imposes the unit root constraints that $\Phi(1) = I$.
2. Unless there exists "sufficient cointegration with respect to the variable(s)" from which causality is investigated, Granger non-causality tests have nonstandard asymptotic distributions that are function of nuisance parameters.

A more precise and practical meaning of "sufficient cointegration" will be given in the following section as it is more easily cast in the context of a vector error-correction model. Nonetheless, the following special cases illustrate what issues are involved to justify special cases for which the asymptotic standard distributions of Granger non-causality tests.

1. Sims, Stock and Watson (1990) have shown that in a bivariate *VAR* model with $I(1)$ variables, Granger non-causality tests follow the usual distributions if both variables are cointegrated.
2. In a trivariate *VAR* model, say with x_{1t}, x_{2t}, and x_{3t} being each $I(1)$, they have also shown that a test of Granger non-causality of x_{2t} on x_{1t} will

follow the usual distribution if x_{2t} is cointegrated with either x_{1t} and/or x_{3t}.

3. In the trivariate model, Toda and Phillips (1994) show that even if there is no cointegration in the system, a non-rejection of Granger non-causality based on the levels *VAR* with conventional critical values is still valid. In this case, that is because the true distribution of the test for a given significance level, involves convergence of the statistic to values that are larger than that of standard distributions. Note that a rejection of Granger non-causality is uninformative.[52]

4. Still, if we want to test that the block of variables defined by x_{2t} and x_{3t} is not Granger causing x_{1t}, then Toda and Phillips (1993b, 1994) have proved that two cointegrating vectors involving x_{2t} and x_{3t} are required to validate the use of standard distributions.

Inference in levels *VAR* models with integrated regressors requires much attention. Except for some of the cases we have discussed, other test statistics will have unit root distributions which may asks to generate appropriate critical values from Monte Carlo simulations. Other avenues include the use of fully modified least-squares estimation of the levels *VAR*, developped by Phillips (1993c), and Johansen's (1988) system maximum likelihood estimation.[53] The latter is the object of the next section.

4.4.3 Vector error-correction models (VECM) and cointegration in dynamic equation systems

4.4.3.1 The set-up and its usefulness

In a way similar to the rewriting of an *ADL* model into an *ECM* representation, one can transform a *VAR* system in levels into a vector autoregressive error-correction model (*VECM*). Johansen (1988), and Johansen and Juselius (1990, 1992) have developed econometric tools specific to this framework.[54] If we consider a slight generalization of the *VAR* system given by equation (4.44), by allowing a deterministic trend and some $s \times 1$ vector S_t of exogenous $I(0)$ regressors to be added to the system, we have:

$$\mathrm{X}_t = \mu_0 + \mu_1 t + \Phi_1 \mathrm{X}_{t-1} + \Phi_2 \mathrm{X}_{t-2} + \cdots + \Phi_p \mathrm{X}_{t-p} + \Upsilon \mathrm{S}_t + \varepsilon_t , \tag{4.75}$$

---

[52] Remember that with no cointegration, the Granger non-causality tests based on the differences *VAR* would follow standard distributions.

[53] Another possibility, which we have not explored in this guide is the application of Bayesian econometrics that is conditional on initial observations and Gaussian errors, for which the likelihood function does not need to be different with stationary or nonstationary variables. This possibility is evoked by Sims, Stock and Watson (1990) in their conclusions.

[54] Although these tools have recently been extended to deal with the case of $I(2)$ variables, e.g. Johansen (1991b, 1992b), our discussion will focus on systems with $I(1)$ variables.

for $t = 1, 2, ..., T$, where $\mathbf{X}_t$, μ_0 and μ_1 are $N \times 1$, Φ_i is $N \times N$, for $i = 1,\dots,p$, and Υ is $N \times s$.[55] We also assume that p is known, that $\mathbf{X}_{-p+1}$, ..., and $\mathbf{X}_0$ are fixed, and that ε_t is a normally distributed vector white noise process.

By adding and subtracting $\Sigma_{j=1}^{p}\left[\left(\Sigma_{i=1}^{j}\Phi_{i-1}\right)\mathbf{X}_{t-j}\right]$, with Φ_0=I, on the RHS of equation (4.75), the *VAR*(*p*) system can be reparameterized in first-difference form, so that we get:[56]

$$\Delta\mathbf{X}_t = \mu_0 + \mu_1 t + \Gamma_1 \Delta\mathbf{X}_{t-1} + \cdots + \Gamma_{p-1}\Delta\mathbf{X}_{t-p+1} + \Pi\mathbf{X}_{t-p} + \Upsilon\mathbf{S}_t + \varepsilon_t, \tag{4.76}$$

where the $N \times N$ matrices of coefficients are defined as followed:

$$\Gamma_i = -(\mathbf{I}_N - \Phi_1 - \cdots - \Phi_i) \text{ for } i = 1,\dots,p-1, \tag{4.77a}$$

$$\Pi = -(\mathbf{I}_N - \Phi_1 - \cdots - \Phi_p). \tag{4.77b}$$

Johansen's method aims at analyzing the properties of matrix Π regarding information about long-run relationships between elements of $\mathbf{X}_t$. Reformulating this $N \times N$ matrix as the product of two $N \times r$ full-rank matrices, for some $r \leq N$ to be determined, i.e.

$$\Pi = \alpha\beta', \tag{4.78a}$$

or

$$\Pi = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1N} \\ \vdots & \ddots & \vdots \\ \pi_{N1} & \cdots & \pi_{NN} \end{bmatrix}_{N \times N} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1r} \\ \vdots & \ddots & \vdots \\ \alpha_{N1} & \cdots & \alpha_{Nr} \end{bmatrix}_{N \times r} \begin{bmatrix} \beta_{11} & \cdots & \beta_{r1} \\ \vdots & \ddots & \vdots \\ \beta_{1r} & \cdots & \beta_{rN} \end{bmatrix}_{r \times N}. \tag{4.78b}$$

Written in this format, we can see more easily why equation (4.76) is a *VECM*. The matrix β is made of r cointegrating vectors. Hence, the elements of $\beta'\mathbf{X}_t$ are the error-correction terms. The elements of matrix α are the adjustment coefficients associated with the error-correction terms. Thus, each equation of the *VECM* is an error-correction model, with possibly more than one error-correction term if there are multiple cointegrating vectors between the endogenous variables of the system. In this framework, we can also see why some error-correction terms may appear in more than one equation. This leads therefore to the situation for which the elements of X_t are not weakly exogenous with respect to the elements of the corresponding row of β. Efficiency gains must therefore exist from estimating jointly the system of equations given by (4.76).

---

[55] For example, the matrix of exogenous stationary regressors could be made up of centered seasonal dummy variables, i.e. that each averages zero over the sample. This particular definition of seasonal dummies is crucial so that they would not imply seasonal drifts in the level of the series. For instance with quarterly data, three seasonal dummies, from *i=1, 2,* or *3* could be defined as: $D_i = 1$ for the i^{th} quarter, -1 for the fourth quarter, and 0 otherwise, as long as each one sums to zero over the sample.

[56] It is also possible to write a *VECM* reparameterisation of the *VAR* such that matrix Π will be associated with $\mathbf{X}_{t-1}$. The other matrices of coefficients will then be defined somewhat differently. This can easily be shown by subtracting $\mathbf{X}_{t-1}$ from both sides of equation (4.71), so that $\Pi = -\mathbf{I} + \Xi$. This is the format employed by Eviews 1.0.

From the restriction given by equation (4.78a), the rank of matrix Π has to be equal to r, i.e. the number of independent cointegrating relationships there exists between the N variables of the system. There are three possibilities:

1. If $\text{rank}(\Pi) = N$, so that the Π has full rank, all elements in X_t are stationary;
2. If $\text{rank}(\Pi) = 0$, so that Π is the null matrix, there are no cointegrating relationships between the variables in X_t -- the system of equations given by (4.53) is a *VAR* for first-differences of the series;
3. If $0 < \text{rank}(\Pi) = r < N$, there exists two $N \times r$ matrices α and β such that $\Pi = \alpha\beta'$ and $\beta' X_t \sim I(0)$ even if $X_t \sim I(1)$.

Johansen's procedure will therefore consist in estimating by full information ML the system in (4.76) subject to restriction (4.78a).[57] Then, it is possible to construct statistics, like LR tests, to determine the rank of Π, or to test restrictions on the intercept, on the coefficient of the deterministic trend, or on the elements in the α and β matrices. Before entertaining these econometric issues, we must discuss other questions regarding the set-up.

First, we must notice that while the number of independent vectors forming the β' matrix is unique, the matrices α and β are not uniquely determined without further restrictions (except if $N = 2$ and $r = 1$) . Indeed, for any non-singular matrix C, we have that $\Pi = \alpha\beta' = \alpha C$ $C^{-1} \beta'$. This is why if there is more than one cointegrating vector, estimates of the columns of β are said to provide information about the cointegrating space spanned by these vectors. However, as discussed in Phillips (1991b), without prior identifying restrictions, the individual coefficients of β should not be interpreted when $r > 1$. In many instances, provided that the coefficients associated with the specific variables are non zero, one can normalize to unity one element in each row of β'. Nonetheless, these r restrictions are not sufficient to identify the structural cointegrating vectors. Sometimes economic theory will suggest additional identifying restrictions of β'. In other situations, Phillips (1991b, 1994) has recommended a "triangular system representation [that] explicitly incorporates identifying conditions on the cointegration space that isolate a set of full rank integrated regressors and identify the cointegrating equations as individual structural equations." (p. 73, Phillips, 1994).

---

[57] This is also called the technique of reduced rank regression.

This triangular identification scheme involves partitioning the β' matrix in two submatrices. Let us suppose that some initial orderings of the variables in X_t imply some values for β' whose rows form a basis of the cointegrating space. Provided that the element of the first row and first column of that matrix is nonzero, it is possible to normalise it to unity with respect to this variable. If not, a reordering of the variables in X_t and therefore of the rows in β' may ensure that the first variable in the new ordering belongs to a cointegrating relation. The same idea can be carried out for each of the r cointegrating relations. Then, given this ordering, simple operations on the rows of β' will imply that the rows of the following matrix are linearly independent and that they also span the cointegrating space. That is,[58]

$$\beta' = \begin{bmatrix} 1 & 0 & \cdots & 0 & -b_{1r+1} & -b_{1r+2} & \cdots & -b_{1N} \\ 0 & 1 & \cdots & 0 & -b_{2r+1} & -b_{2r+2} & \cdots & -b_{2N} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & -b_{r\,r+1} & -b_{r\,r+2} & \cdots & -b_{rN} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_r & -\mathbf{B}' \end{bmatrix}, \tag{4.79}$$

where $\mathbf{I}$ is $r \times r$ and $\mathbf{B}'$ is $r \times (N-r)$. Let us define conformably X_{1t} and X_{2t} with dimensions $r \times 1$ and $(N-r) \times 1$ respectively. Given that $\Delta X \sim I(0)$, in the most general case, we can define $E(\Delta X) = \theta_0 + \theta_1\, t$, so that

$$\Delta X_t = \begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \begin{bmatrix} \theta_{01} \\ \theta_{02} \end{bmatrix} + \begin{bmatrix} \theta_{11} \\ \theta_{12} \end{bmatrix} t + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{4.80}$$

Also, in a general case, we can think of $\beta' X_t$ as generating a stationary process z_t around a linear determinitic trend, i.e.

$$\beta' X_t = [\mathbf{I}_r \;\; -\mathbf{B}'] \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \delta_0 + \delta_1\, t + z_t. \tag{4.81}$$

---

[58] As illustrated in Hamilton (1994), these operations will be like multiplying the first row of β' by the (2,1) element of the original matrix and subtracting it from the second row, multiplying the first row of β' by the (3,1) element of the original matrix and subtracting it from the third row, etc. until the first column of β' be equal to the first column of $\mathbf{I}_r$. We would then go on similarly with respec to the other columns of β' until we obtain the partitioning in equation (4.79).

Using equations (4.80) and (4.81), we can then write Phillips' (1991b, 1994) triangular representation of a cointegrating system in a general case as:

$$\mathbf{X}_{1t} = \delta_0 + \delta_1 t + \mathbf{B}'\mathbf{X}_{2t} + \mathbf{z}_t, \tag{4.82a}$$

$$\Delta\mathbf{X}_{2t} = \theta_{02} + \theta_{12} t + \mathbf{u}_{2t}. \tag{4.82b}$$

This representation is equivalent to that of the *VECM* of equation (4.76), except that it imposes identifying restrictions on the cointegrating vectors given by equation (4.79).[59]

Second, the intercept and the deterministic trend in equation (4.76) may imply different things about the data generating process depending whether some restrictions on these coefficients are valid. Let us define the following matrices: $\mathbf{P}_\alpha \equiv [\alpha(\alpha'\alpha)^{-1}\alpha']$ and $\mathbf{D}_\alpha \equiv [(\alpha'\alpha)^{-1}\alpha']$. Adding and subtracting $[\mathbf{P}_\alpha\,\mu_0 + \mathbf{P}_\alpha\,\mu_1\, t]$ to equation (4.76), then regrouping some of the terms, we obtain:

$$\begin{aligned}\Delta\mathbf{X}_t &= [\mathbf{I} - \mathbf{P}_\alpha]\,\mu_0 + [\mathbf{I} - \mathbf{P}_\alpha]\,\mu_1 t + \Gamma_1\,\Delta\mathbf{X}_{t-1} + \cdots \\ &\quad + \Gamma_{p-1}\,\Delta\mathbf{X}_{t-p+1} + \alpha\left[\beta'\mathbf{X}_{t-p} + \mathbf{D}_\alpha\,\mu_0 + \mathbf{D}_\alpha\,\mu_1 t\right] + \Upsilon\mathbf{S}_t + \varepsilon_t.\end{aligned} \tag{4.83}$$

If we define the $N \times (N\text{-}r)$ orthogonal complement matrix $\alpha_\perp$ of α, that is such that $\alpha_\perp'\alpha = \alpha'\alpha_\perp = 0$, and defining similarly $\mathbf{P}_{\alpha\perp}$, and $\mathbf{D}_{\alpha\perp}$ as being respectively the orthogonal complement of $\mathbf{P}_\alpha$ and $\mathbf{D}_\alpha$ we can rewrite equation (4.83) as:

$$\begin{aligned}\Delta\mathbf{X}_t &= \mathbf{P}_{\alpha\perp}\,\mu_0 + \mathbf{P}_{\alpha\perp}\,\mu_1 t + \Gamma_1\,\Delta\mathbf{X}_{t-1} + \cdots \\ &\quad + \Gamma_{p-1}\,\Delta\mathbf{X}_{t-p+1} + \alpha\left[\beta'\mathbf{X}_{t-p} + \mathbf{D}_\alpha\,\mu_0 + \mathbf{D}_\alpha\,\mu_1 t\right] + \Upsilon\mathbf{S}_t + \varepsilon_t.\end{aligned} \tag{4.84}$$

This can also be rewritten as:

$$\begin{aligned}\Delta\mathbf{X}_t &= \alpha_\perp\theta_0 + \alpha_\perp\theta_1 t + \Gamma_1\,\Delta\mathbf{X}_{t-1} + \cdots \\ &\quad + \Gamma_{p-1}\,\Delta\mathbf{X}_{t-p+1} + \alpha\left[\beta'\mathbf{X}_{t-p} + \delta_0 + \delta_1 t\right] + \Upsilon\mathbf{S}_t + \varepsilon_t\end{aligned} \tag{4.85a}$$

where $\theta_0 \equiv \mathbf{D}_{\alpha\perp}\,\mu_0,$ (4.85b)

$$\theta_1 \equiv \mathbf{D}_{\alpha\perp}\,\mu_1, \tag{4.85c}$$

$$\delta_0 \equiv \mathbf{D}_\alpha\,\mu_0, \tag{4.85d}$$

---

[59] The same identifying restrictions can also be imposed in the *VECM* representation. That is what is done when applying the Johansen estimation procedure in Eviews 1.0.

$$\delta_1 \equiv D_\alpha \mu_1 . \tag{4.85e}$$

Comparing equations (4.76) and (4.85a), we see that the unrestricted values of the intercept and of the coefficient of the deterministic time trend each have possibly two components:

$$\mu_0 = \alpha\delta_0 + \alpha_\perp\theta_0 , \tag{4.86}$$

$$\mu_1 = \alpha\delta_1 + \alpha_\perp\theta_1 . \tag{4.87}$$

Hence, as documented by Osterwald-Lenum (1992), there are several possibilities, which we summarize in Table 4.7. Cases 1 to 5 show the passage from the most general to the most specific representation of the data generating process.

1. In the most general case for which both μ_0 and μ_1 would be nonzero and unrestricted, the DGP for the vector X_t would in effect contain a quadratic polynomial in time. The model specializes for the value of μ_1 being defined from $\theta_1 \neq 0$.

2. It is possible however that the value of μ_1 be contained in the space spanned by α, so that $\mu_1 = \alpha\delta_1$ and $\theta_1 = 0$. Equivalently, this means that $\alpha_\perp'\mu_1 = 0$ so that the long-run relationship evolves around a deterministic linear trend.

3. A more specialized specification with $\mu_1 = 0$ would involve a linear trend in the level of the series given by $\alpha_\perp'\theta_0$, and a drift given by $\alpha\delta_0$.

4. If, furthermore, $\theta_0 = 0$, i.e. $\alpha_\perp'\mu_0 = 0$, but if $\delta_0 \neq 0$ so that $\mu_0 = \alpha\delta_0$, the error-correction terms are zero-mean stationary processes. There is no drift in the first-difference of X_t, and there is no linear trend in its level.

5. Finally, the most restricted specification of the DGP occurs when both δ_0 and θ_0 are zero, so that $\mu_0 = 0$. In this case, no intercept would be included in the model.

In practical work, cases 3 and 4 have been more frequently considered. This is why we will refer to these cases in the next section.

Having discussed the set-up of a simultaneous *ECM* equations system, we now turn to the issues specific to estimation and inference on its parameters.

Table 4.7.- Restricted and unrestricted intercept, and deterministic trend in a *VECM*.

DGP model in levels (from equation (4.75)):[1]

$$\Phi(L)\, X_t = \mu_0 + \mu_1 t + \Upsilon S_t + \varepsilon_t$$

$$\Rightarrow (1-L)\,\Phi^*(L)\, X_t = \mu_0 + \mu_1 t + \Upsilon S_t + \varepsilon_t$$

$$\Rightarrow X_t = \left(\Phi^*(L)^{-1}\right)\mu_0\, t + \left(\Phi^*(L)^{-1}\right)\mu_1 (t^2 - t) + \Phi^*(L)^{-1}(1-L)^{-1}\Upsilon S_t + \Phi^*(L)^{-1}(1-L)^{-1}\varepsilon_t$$

DGP model in first differences:

$$\Delta X_t = \mu_0 + \mu_1 t + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1}\Delta X_{t-p+1} + \alpha\beta' X_{t-p} + \Upsilon S_t + \varepsilon_t$$

The statistical model:

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1}\Delta X_{t-p+1} + W_t + \Upsilon S_t + \varepsilon_t$$

Cases:	Deterministic parts in two representations of DGP					Restrictions on the DGP		Specification of the statistical model $W_t =$
	Levels			Differences				
	Intercept	Linear trend	Quadratic trend	Drift	Linear trend			
Case 1	yes	yes	restricted	yes	yes		$\delta_1, \theta_1 \neq 0$	$\alpha\beta' X_{t-p} + \mu_0 + \mu_1 t$
Case 2	yes	yes	no	yes	no	$\theta_1 = 0$	$\delta_1 \neq 0$	$\alpha\begin{pmatrix}\beta' & \delta_1\end{pmatrix}\begin{pmatrix}X_{t-p}\\ t\end{pmatrix} + \mu_1$
Case 3	yes	restricted	no	yes	no	$\mu_1 = 0$	$\delta_0, \theta_0 \neq 0$	$\alpha\beta' X_{t-p} + \mu_0$
Case 4	yes	no	no	no	no	$\theta_0, \mu_1 = 0$	$\delta_0 \neq 0$	$\alpha\begin{pmatrix}\beta' & \delta_0\end{pmatrix}\begin{pmatrix}X_{t-p}\\ 1\end{pmatrix}$
Case 5	no	no	no	no	no	$\mu_0, \mu_1 = 0$		$\alpha\beta' X_{t-p}$

[1] $\Phi(L) \equiv (I - \Phi_1 L - \ldots - \Phi_p L^p) \equiv (1-L)(I - \Phi_1^* L - \ldots - \Phi_p^* L^p) \equiv (1-L)\,\Phi^*(L)$ where $\Phi^*(L)$ is stationary.

4.4.3.2 Johansen's estimation and testing procedures

4.4.3.2.1 FIML estimation of a *VECM*

Let us consider the following *VECM* (imposing $\mu_1 = 0$ at the outset to simplify the exposition) over an effective sample of T observations,[60]

$$\Delta X_t = \mu_0 + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1}\Delta X_{t-p+1} + \Pi X_{t-p} + \Upsilon S_t + \varepsilon_t, \qquad (4.88)$$

that we want to estimate by full information ML subject to the restriction that the matrix Π has a reduced rank r, i.e.

$$H_2(r):\ \Pi = \alpha\beta' \qquad \text{and} \qquad rank(\Pi) = r. \qquad (4.89)$$

---

[60] It is assumed that p first observations of X_u have already been set aside to define the lagged dependent variables in the system.

Johansen (1988, 1992a) has shown that provided that the errors are normally distributed for a given lag-length p, we can obtain the ML estimates of μ_0, Γ_i (for $i = 1, ..., p-1$), α, β and Σ_ε following the steps below.[61]

1. We purge ΔX_t and X_{t-p} from the influence of the dynamics and of the intercept (if left unrestricted, or as in case 3). This is done by constructing two $N \times 1$ vectors R_{0t} and R_{pt}, as the vectors of residuals from the following OLS regressions:

$$\Delta X_t = \mu_0 + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-p+1} + \Upsilon S_t + R_{0t}, \tag{4.90a}$$

$$X_{t-p} = \mu_0 + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-p+1} + \Upsilon S_t + R_{pt}. \tag{4.90b}$$

If, however, the intercept is restricted to enter the cointegrating relationship only (as in case 4), then R_{pt} is a $(N+1) \times 1$ vector of residuals obtained from the following OLS regression:

$$\begin{pmatrix} X_{t-p} \\ 1 \end{pmatrix} = \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-p+1} + \Upsilon S_t + R_{pt}. \tag{4.90b'}$$

2. If we consider a model with an unrestricted intercept, or a specification corresponding to case 3, the next step concerns the estimation of α, β and Σ_ε in

$$R_{0t} = \alpha\beta' R_{pt} + \varepsilon_t. \tag{4.91}$$

For the specification of case 4, the model to be considered is given by:

$$R_{0t} = \alpha \begin{pmatrix} \beta' \\ \delta_0' \end{pmatrix} R_{pt} + \varepsilon_t. \tag{4.91'}$$

The log-likelihood function to be maximized for estimating equation (4.91) is given by:

$$\log l(\alpha, \beta, \Sigma_\varepsilon) = \text{constant} - \frac{T}{2} \log |\Sigma_\varepsilon| - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t' \Sigma_\varepsilon \varepsilon_t . \tag{4.92}$$

(The problem is similar for equation (4.91')).

Johansen (1988) has shown that the maximization of this function is equivalent to solving the following eigenvalue problem for its r largest roots:

$$\left| \tau S_{pp} - S_{p0} S_{00}^{-1} S_{0p} \right| = 0, \tag{4.93}$$

where the residual product moment matrices are defined as:

---

[61] This presentation does not intend to review all the mathematical and statistical foundations of the procedure. This is done in the original papers, as well as in Banerjee *et al.* (1993). Rather, it intends to give a heuristic feel for the estimation procedure.

$$\mathbf{S}_{ij} = T^{-1}\mathbf{R}_{ij}'\mathbf{R}_{ij}, \qquad \text{for} \qquad i,j = 0,p. \tag{4.94}$$

Then, ranking the N eigenvalues that solve equation (4.93) in descending order, i.e. $\hat{\tau}_1 > \hat{\tau}_2 > \cdots > \hat{\tau}_N$, and denoting the N orthonormal corresponding eigenvectors by $\hat{\mathbf{V}} = (\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_N)$, the ML estimate of β is given by the eigenvectors corresponding to the r largest eigenvalues:

$$\hat{\boldsymbol{\beta}} = (\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_r),$$

and the ML estimation of α is given by:

$$\hat{\boldsymbol{\alpha}} = \mathbf{S}_{0p}\hat{\boldsymbol{\beta}},$$

so that the reached maximum of the log-likelihood is given by:

$$\log l(\alpha, \beta, \Sigma_\varepsilon) = -\frac{NT}{2} - \frac{NT}{2}\log 2\pi - \frac{T}{2}\log|\mathbf{S}_{00}| - \frac{T}{2}\sum_{i=1}^{r}\log(1-\hat{\tau}_i). \tag{4.95}$$

An important caveat must be emphasized regarding the estimates of β. As shown by Phillips (1991b), without prior identifying restrictions, the point estimates of the elements of β converge to random variables not constants, and individual coefficients should not be interpreted with any specific meaning or tested for any specific values when $r > 1$.[62] The exception to this comment is when $r = 1$, so that a single normalisation with respect to one variable makes the interpretation straightforward and can be sometimes linked to some economic theory. When $r > 1$, without prior identifying restrictions, only hypothesis testing with respect to the cointegrating space can be performed, and even then, the non-rejection of the null hypothesis should be taken only as being possibly consistent with a theoretical interpretation. The individual coefficients of the structural cointegrating vectors are still unidentified. The alternative is to specify at the outset a sufficient number of identifying restrictions prior to the estimation. Phillips' (1991b, 1994) triangular representation is one way to ensure identification so that the individual structural equations are consistently estimated.

4.4.3.2.2 Determination of the cointegrating rank of a *VECM*

Once the estimation of a *VECM* is carried out under the unrestricted hypothesis that there are up to N cointegrating relationships, the N eigenvalues that solve equation (4.93) can be used to construct two versions of *LR*-statistics to test down the cointegrating rank of the long-run matrix Π.

---

[62] Hodgson (1992) presented some very illustrative simulations of the convergence of the ML estimator of β' to random variables.

If we wish to test the null hypothesis

$$H_{0,2}(r): rank(\Pi) = r \leq r_1 \in [0, N[\tag{4.96a}$$

against the alternative

$$H_{1,2}(r): rank(\Pi) = r = r_1 + 1 \in [r_1, N[, \tag{4.96b}$$

Johansen has proposed the maximal eigenvalue statistic defined as:

$$MV(r_1) = -T\log(1 - \hat{\tau}_{r_1+1}) \tag{4.97}$$

where $\hat{\tau}_{r_1+1}$ is the $(r_1 + 1)^{th}$ largest eigenvalue found at the estimation stage. This statistic follows a non-standard asymptotic distribution tabulated by Osterwald-Lenum (1992), with $(N - r_I)$ degrees of freedom.

Let us now consider the testing of

$$H_{0,2}(r): rank(\Pi) = r = r_1 \in [0, N[\tag{4.98a}$$

against

$$H_{1,2}(r): rank(\Pi) = r \geq r_1 + 1, \text{ i.e. } r \in]r_1, N]. \tag{4.98b}$$

Then, we can test the null hypothesis in equation (4.96a) with the following trace statistic:

$$TR(r_1) = -T\sum_{i=r1+1}^{N} \log(1 - \hat{\tau}_i) \tag{4.99}$$

where the $\hat{\tau}_i$'s are the $(N - r_1 - 1)^{th}$ smallest eigenvalues computed at the estimation stage. This test statistic follows another non-standard asymptotic distribution tabulated by Osterwald-Lenum (1992), with $(N - r_I)$ degrees of freedom. In practice, the test is conducted in a sequential manner, varying r_I from 0 to N-1.

While performing these tests is fairly straightforward once we have computed the eigenvalues of the system, some caveats must be known to obtain and to interpret appropriately their results.

First, the distribution of either $MV(r_1)$ or $TR(r_1)$ depends on how the intercept μ_0 has been dealt with in the statistical representation being estimated. If case 3 applies and a restricted linear trend enters the DGP in levels, while a drift is present in the differences of the DGP, following the estimation of a corresponding statistical model, table A4.2 should be consulted. Now, suppose that case 4 applies with a common mean in the error-correction term while there is no linear trend in the DGP of vector $\mathbf{X}_t$, and that we estimate the corresponding statistical model, the asymptotic distribution of the *MV* and *TR* statistics are reported in table A4.3. Finally, if case 4 applies, but that we nevertheless estimate a statistical model with an unrestricted intercept, the appropriate critical values for the cointegrating rank should be retrieved from table A4.4.[63]

Second, Pitarakis (1993) has shown that the *MV* and *TR* tests can have poor performances in finite sample when compared to their asymptotic critical values. In many instances, the problem will be the power of the tests so that one would tend to find too often no cointegration. But, for large values of N relative to T, oversized tests may make it hard not to find wrongly at least one cointegrating vector. He suggests instead to use either Reinsel and Ahn (1988) corrected version of the *TR*-statistic, given by

$$RATR = -(T-Np)\sum_{i=r+1}^{N}\log(1-\hat{\tau}_i), \tag{4.100}$$

the Pitarakis-Gonzalo (*PG*) statistic, defined as

$$PG = \frac{T}{2}\left(\sum_{i=r+1}^{N}\hat{\tau}_i - \sum_{i=r+1}^{N}\log(1-\hat{\tau}_i)\right), \tag{4.101}$$

or the corrected Pitarakis-Gonzalo (*PGp*) statistic, defined as

$$PGp = \frac{1}{2}\left((T-Np+p)\sum_{i=r+1}^{N}\hat{\tau}_i - (T-Np+N/2)\sum_{i=r+1}^{N}\log(1-\hat{\tau}_i)\right). \tag{4.102}$$

The estimated eigenvalues are obtained at the estimation stage, and are ranked in dscending order. His simulations suggest that the critical values of both statistics remain stable for different sample sizes when $N - r \leq 5$, and that, for $N - r > 5$, the distribution of the *PG* statistic remains closer to the asymptotic distribution. The latter is true for the *PGp*-statistic with both small and large values of *N*.[64]

---

[63] Readers, who would be considering applications with DGP and statistical models of cases 1 and 2, can find the asymptotic distribution of the cointegrating rank tests in Osterwald-Lenum (1992).

Another issue concerns the correct determination of the lag length *p* since it is assumed to be known at the outset when applying Johansen's procedure. As shown by Pitarakis (1993), this is not a trivial issue since both under- and over-parameterization of the system will distort inferences about cointegration. If *p* is chosen to be too small, the omitted dynamics should generally result in biases and inconsistencies in the ML estimates of the model parameters. In this instance, serially correlated errors should be symptomatic of the problem. But also, if *p* is chosen to be too large, the finite sample critical values of the conventional *MV*- and *TR*-statistics are much bigger, so that overfitting will usually lead these tests to find cointegration too often. Distortions in inferences by overfitting will also be a bigger problem in large system. Pitarakis (1993) therefore recommends the use of penalized versions of the statistics such as *RATR*, *PG*, and especially *PGp*.

Johansen and Juselius (1990, 1992) select their lag-length *p* so that the residuals of each equation in the system (4.75) (or equivalently in (4.76)) are normally distributed white noise, insofar as these hypotheses are not rejected at conventional statistical significance levels by *LM*-tests for serial correlation and the Bera-Jarque test. An alternative lag-length selection procedure could make use of information criteria. Reimers's (1991) simulation exercises suggested that Schwarz information criterion did perform reasonably well. However, since this criterion attributes a big penalty for extra lags, it will favour a smaller lag length, which can lead to underfitting problem.

Finally, to feel confident about the number of cointegrating vectors, Johansen and Juselius (1992) recommend to examine two sets of plots. The first set is made of graphs of the residual vectors associated with the *r* cointegrating vectors, i.e. $\beta_i' X_t$, where β_i is the i^{th} column of β, for $i = 1, ..., r$. These graphs illustrate the short-run actual deviations from long-run equilibria between the integrated variables. Also, the cointegrating relationships associated with the largest eigenvalues of the system are going to be the most correlated with the stationary part of the process ΔX_t corrected for the short-run dynamics. Even if there exist long-run or cointegrated relationships between non-stationary series, innovations and their short-run dynamics in a sample may keep these variables from reaching their long-run equilibrium. This is why a second set of graphs that plot the residuals of the cointegrating vectors adjusted fo the short-run dynamics should be examined as well. These are defined as $\beta_i' R_{pt}$, where β_i is the i^{th} column of the basis generating the cointegrating space, and R_{pt} is defined as a matrix of the residual vectors obtained from estimating equation (4.90b). These graphs should illustrate over the sample a stationary adjustment path corrected for the short-run dynamics of the model. The more complex and rich is the short-run dynamics of the model, the more different the two sets of graphs will appear. As explained by Johansen and Juselius (1992), "*an important property of this* [vector error-correction] *model* [is] *its ability to describe an inherent tendency to move towards the* [long-run] *equilibrium states, without necessarily ever reaching it because of frequent and often large shocks pushing it away from the equilibrium path.*"

---

[64] Pitarakis (1993) also proposed to use an information criterion to select the cointegrating rank *r*. He studied the performance of both Akaike and Schwarz type criteria to select the cointegrating rank for a given lag length. His simulations suggested that Akaike's criterion for the cointegrating rank leads too often to find a positive rank when the true *r* is effectively zero, even though it tends to find the right cointegrating rank if the true *r* is greater than zero. By comparison, Schwarz information criterion found cointegrating rank that is biased toward choosing a very small rank. After simulations, he recommended the use of a *ICPG* criterion, defined as $ICPG = PG - 3.1\,(N - r)^2$.

4.4.3.2.3 Hypothesis testing about the presence or absence of a linear trend

As discussed previously, the critical values of the cointegrating rank tests are dependent of the treatment of the deterministic trend or of the intercept in the *VECM*. It is always possible to refer to the largest critical values, which amounts to use the distribution corresponding to the most restrictive specification. In this case, one will make it harder to find a high cointegrating rank. Another way to decide to include either an intercept in the cointegrating relationship itself (case 4 in table 4.7), or to allow a drift in the DGP of the first-difference series (case 3 in table 4.7), is to examine visually plots of the data. Remember that the latter instance corresponds to the presence of a restricted deterministic time trend in the level of the data.

Another possibility (in itself or as a way to validate or not the conclusion based on the visual inspection) is to test directly the null hypothesis of no drift in the differences of the DGP. This can be done following the steps described below.

1. Estimate the specification of the statistical model under the hypothesis that there is a linear trend in the levels of the data. This will be the unrestricted model. Its specification is given by the last column in table 4.7 for case 3. Test for the cointegrating rank *r*. Get the *N-r* smallest eigenvalues under this hypothesis: $\hat{\tau}_{r+1}, \ldots, \hat{\tau}_N$.

2. Estimate the specification of the statistical model under the hypothesis that the intercept enters only through the cointegrating relationship. This is the restricted model. The relevant specification is given by the last column in table 4.7 for case 4. Test for the cointegrating rank *r*. Get the *N-r* smallest eigenvalues under this hypothesis: $\hat{\tau}^*_{r+1}, \ldots, \hat{\tau}^*_N$. (Notice that you will have to decide of a common value for the cointegrating rank *r* to perfom this step).
3. To test H_0: case 4: $\theta_0 = 0$ against H_1: case 3: $\theta_0 \neq 0$, use the likelihood ratio statistic given by:

$$LR_{\theta_0} = -T \sum_{i=r+1}^{N} \log\left[\frac{1-\hat{\tau}^*_i}{1-\hat{\tau}_i}\right] \overset{a}{\sim} \chi^2(N-r). \tag{4.103}$$

Extensions of this procedure to test case 1 against case 2, or case 2 against case 3 are straightforward.

4.4.3.2.4 Hypothesis testing about the adjustment matrix α

We discussed in section 4.3.3.1.3 the role of weak exogeneity of the integrated variables for α and β to justify efficiency of some individual equation estimation procedures of the cointegrating coefficients, such as the GIVE estimation of a Bewley representation of an ECM, or the direct NLS estimation of the ECM. Otherwise the use of optimal single equation estimators (discussed in section 4.3.3.1.4) or estimation of a system of equations is warranted. In the same vein, one could limit oneself in considering only a partial system if he or she were confident that the other integrated variables of a fuller system were weakly exogenous. Johansen and Juselius (1990), and Johansen (1992a) proposed to formulate a test for weak exogeneity of specific parameters of interest as a test on the relevant elements of the α matrix. Weak exogeneity of some integrated variable X_{it} for a set of long-run parameters in a given equation of the system demands that the error correction terms involving these parameters do not belong in some other equations determining X_{it}. Before stating the general result, let us illustrate the argument with a specific example.

Let us consider the case of a system of $N = 3$ integrated variables that are related by $r = 2$ cointegrating relationships. This would imply in the most general system that the Π matrix, whose rank is 2, be such that:

$$\Pi\,X_{t-p} = \underset{3\times 2}{\alpha}\;\underset{2\times 3}{\beta'}\;\underset{3\times 1}{X_{t-p}} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} X_{1t-p} \\ X_{2t-p} \\ X_{3t-p} \end{bmatrix} . \tag{4.104a}$$

$$= \alpha\,(\beta' X_{t-p}) = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} X_{1t-p} + \beta_{21} X_{2t-p} + \beta_{31} X_{3t-p} \\ \beta_{12} X_{1t-p} + \beta_{22} X_{2t-p} + \beta_{32} X_{3t-p} \end{bmatrix} . \tag{4.104b}$$

Hence, in the first equation of the system, weak exogeneity of both X_{2t} and X_{3t} for all the parameters in β can be associated with the following restricted α matrix whose second and third row have zero elements, i.e.

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{3\times 1} \begin{bmatrix} \psi_{11} & \psi_{12} \end{bmatrix}_{1\times 2} = A\psi , \tag{4.105}$$

where A is a matrix of known constants reflecting the hypothesized restrictions, and ψ is a matrix of unknown coefficients. If this holds, then OLS estimation of the individual first equation of the system is efficient and equivalent to the FIML estimation of the system.

If we expect the variable X_{3t} to be weakly exogenous with respect to both β_{31} and β_{32}, then both α_{31} and α_{32} must equal zero, i.e.

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}_{3\times 2} \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}_{2\times 2} = A\psi . \tag{4.106}$$

In this case, instead of considering the whole system given by equation (4.88), one could estimate efficiently the partial system made of two equations for X_{1t} and X_{2t} but with $X_{1\,t-p}$, $X_{2\,t-p}$, $X_{3\,t-p}$, lagged values of ΔX_{1t}, ΔX_{2t}, and ΔX_{3t}, Z and an intercept as regressors.

More generally, hypothesis testing on the α matrix can be carried out as a test on the null hypothesis:

$$H_0 :\; \alpha = A_{N\times m}\,\psi_{m\times r} \Leftrightarrow \Pi = (A\,\psi)\,\beta' . \tag{4.107}$$

The estimation of both the unrestricted and the restricted models will provide two sets of r largest eigenvalues given respectively by $\hat{\tau}_1, \ldots, \hat{\tau}_r$, and $\tilde{\tau}_1, \ldots, \tilde{\tau}_r$. The *LR* statistic given by

$$LR_{\alpha} = T \sum_{i=1}^{r} \log\left(\frac{1-\tilde{\tau}_i}{1-\hat{\tau}_i}\right) \overset{a}{\sim} \chi^2(r \cdot (N-m)) \qquad (4.108)$$

can be used to test H_0. For instance, the LR_{α} statistic to test the null hypotheses in equation (4.105) and (4.106) would have (2 (3-1)) = 4 and (2 (3-2)) = 2 degrees of freedom, respectively.

Another way to assess weak exogeneity is to estimate the cointegrating vectors, to construct the cointegrating residuals $\hat{\beta}' X_{t-p}$, then to apply a Wald test about the nullity of the relevant elements of the α matrix in the marginal model equations of the hypothesized weakly exogenous variables. For instance, to test the null hypothesis embedded in equation (4.106), one can estimate

$$\Delta X_{3t} = \mu_{03} + \sum_{i=1}^{p-1} \begin{bmatrix} \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \Delta \mathrm{X}_{t-i} + \begin{bmatrix} \alpha_{31} & \alpha_{32} \end{bmatrix} \beta' \mathrm{X}_{t-p} + \varepsilon_{3t} \qquad (4.109)$$

by OLS and test H_0: $\alpha_{31} = \alpha_{32} = 0$ with a Wald test.

4.4.3.2.5 Hypothesis testing about the basis of the cointegrating space β

We have discussed in section 4.3.3.3, how estimates of the elements of cointegrating vectors are not necessarily asymptotically normally distributed in estimation of single equations. However, as shown by Johansen (1991a), the FIML system analysis provides an estimator of β for which many hypotheses can be tested with statistics that a follow chi-squared distribution. Johansen and Juselius (1992) discussed three types of tests about linear restrictions that could characterise the cointegrating space, given a known cointegrating rank *r*.

First, one may be interested in testing the hypothesis that all the *r* cointegrating vectors are subject to the same *s* homogenous linear restrictions. In this case, the hypothesis is formulated as

$$H_0: \beta = \mathrm{H}_{N \times s}\, \varphi_{s \times r} \quad \Leftrightarrow \quad \Pi = \alpha\, \varphi' \mathrm{H}' \ , \qquad (4.110)$$

where **H** is a matrix of known constants reflecting the common restrictions on all the columns of β, φ is a matrix of unknown elements and $r \le s < N$.[65, 66]

---

[65] Notice that if there is no deterministic time trend in the DGP of the variables in levels, and if an intercept enters the cointegrating relationship, then **H** is a (N+1) × *s* matrix.

[66] Following Phillips (1991b), we have to keep in mind however that these are not hypotheses about the structural cointegrating vectors which are unidentified without prior restrictions. Even though the columns of **H** make explicit the set of vectors that span the cointegrating space, the columns of φ that generate the linear combinations of the cointegrating basis "are still not identified and are subject to arbitrary rotations." (p. 4).

For instance, in a system of $N=3$ variables with $r=2$ independent cointegrating vectors, if one were interested in testing that the first coefficient of β is equal to the second coefficient, this would correspond to

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix} = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix} = \mathbf{H}\varphi \, . \tag{4.111}$$

Also, if one wanted to test that the third component of the cointegrating vector was not different from zero, one would test

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix} = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \\ 0 & 0 \end{bmatrix} = \mathbf{H}\varphi \, . \tag{4.112}$$

As a last example, to test whether the variables into the cointegrating relations have coefficients proportional to (1, -1, -1), the hypothesis would be formulated as

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} \varphi_{11} & \varphi_{12} \end{bmatrix} = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ -\varphi_{11} & -\varphi_{12} \\ -\varphi_{11} & -\varphi_{12} \end{bmatrix} = \mathbf{H}\varphi \, . \tag{4.113}$$

Johansen (1988) has shown that from the estimation of both the unrestricted and the restricted models to obtain two sets of r largest eigenvalues given by $\hat{\tau}_1, \ldots, \hat{\tau}_r$, and $\tau_1, \ldots, \tilde{\tau}_r$, respectively, the following *LR* statistic can be used to test the null hypothesis formulated as (4.110):

$$LR_\beta = T \sum_{i=1}^{r} \log\left(\frac{1-\tilde{\tau}_i}{1-\hat{\tau}_i}\right) \overset{a}{\sim} \chi^2(r \cdot (N-s)) \, . \tag{4.114}$$

For instance, the LR_β statistic to test the null hypotheses in equation (4.111), (4.112), and (4.113) would have (2 (3-2)) = 2, (2 (3-2)) = 2, and (2 (3-1)) = 4 degrees of freedom, respectively.

A second type of linear restrictions on β would consist in fully specifying $r_i \leq r$ of the cointegrating relations that we would treat as known, while leaving the remaining $r - r_i$ cointegrating relations unrestricted. Finally, a third type of linear restrictions on β that may be of interest is that some linear homogenous common restrictions, defined by some **H** matrix of dimension $N \times s$, would apply on r_1 of the cointegrating vectors, while the other $r - r_1$ cointegrating relations would remain unrestricted. Notice that the former type consists in fixing the values of some of the cointegrating vector, but the latter imposes homogenous linear restrictions on a subset of cointegrating vectors without imposing some specific values for their elements. Johansen and Juselius (1992) have developped two *LR* statistic, with $(r_1\,(N - r))$, and $(r_1\,(N - s - (r - r_1))$ degrees of freedom respectively, to test these hypotheses.[67]

4.4.3.2.6 Granger non-causality tests in levels *VAR* and in *VECM*

Section 4.4.2.2 discussed issues regarding Granger non-causality tests in *VAR*'s with integrated variables. Let us consider a $N \times 1$ vector $\mathbf{X}_t$, which includes integrated variables. Let us partition this vector in three blocks $\mathbf{x}_{1t}$, $\mathbf{x}_{2t}$ and $\mathbf{x}_{3t}$, which are respectively of dimension N_1, N_2, and N_3. As a general rule, tests of the hypothesis that $\mathbf{x}_{3t}$ does not Ganger cause $\mathbf{x}_{1t}$ do not asymptotically follow the usual distributions. However, Toda and Phillips (1993b, 1994) have shown that there are exceptions in cases for which there is sufficient cointegration with respect to $\mathbf{x}_{3t}$.[68]

To illustrate the meaning and usefulness of this result, let us rewrite the levels *VAR* and its equivalent *VECM* representation for $\mathbf{X}_t$.[69] The *VAR(p)* representation is given by:

$$\mathbf{X}_t = \Phi(L)\, \mathbf{X}_{t-1} + \mathbf{u}_t, \tag{4.115a}$$

or

$$\begin{bmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \\ \mathbf{x}_{3t} \end{bmatrix} = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} & \phi_{13,1} \\ \phi_{21,1} & \phi_{22,1} & \phi_{23,1} \\ \phi_{31,1} & \phi_{32,1} & \phi_{33,1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1t-1} \\ \mathbf{x}_{2t-1} \\ \mathbf{x}_{3t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \phi_{11,p} & \phi_{12,p} & \phi_{13,p} \\ \phi_{21,p} & \phi_{22,p} & \phi_{23,p} \\ \phi_{31,p} & \phi_{32,p} & \phi_{33,p} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1t-p} \\ \mathbf{x}_{2t-p} \\ \mathbf{x}_{3t-p} \end{bmatrix} + \mathbf{u}_t, \tag{4.115b}$$

and its equivalent *VECM* representation is given by:

$$\Delta \mathbf{X}_t = \Gamma_1\, \Delta \mathbf{X}_{t-1} + \cdots + \Gamma_{p-1}\, \Delta \mathbf{X}_{t-p+1} + \alpha \beta'\, \mathbf{X}_{t-p} + \varepsilon_t, \tag{4.116a}$$

---

[67] Readers are referred to Johansen and Juselius (1992) for a rigorous exposition of these tests and their construction. Pesaran and Pesaran (1991) offer a brief description of these tests in appendix. Micro-FIT 3.0 handles instances of these types of hypotheses.

[68] In effect, these authors have generalised and qualified the results obtained by Sims, Stock, and Watson (1990) for trivariate systems.

[69] To simplify the notation we set μ_0, μ_1 and Υ to zero in equation (4.75).

or

$$\begin{bmatrix} \Delta \mathbf{x}_{1t} \\ \Delta \mathbf{x}_{2t} \\ \Delta \mathbf{x}_{3t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_{11,1} & \boldsymbol{\gamma}_{12,1} & \boldsymbol{\gamma}_{13,1} \\ \boldsymbol{\gamma}_{21,1} & \boldsymbol{\gamma}_{22,1} & \boldsymbol{\gamma}_{23,1} \\ \boldsymbol{\gamma}_{31,1} & \boldsymbol{\gamma}_{32,1} & \boldsymbol{\gamma}_{33,1} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{1t-1} \\ \Delta \mathbf{x}_{2t-1} \\ \Delta \mathbf{x}_{3t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \boldsymbol{\gamma}_{11,p-1} & \boldsymbol{\gamma}_{12,p-1} & \boldsymbol{\gamma}_{13,p-1} \\ \boldsymbol{\gamma}_{21,p-1} & \boldsymbol{\gamma}_{22,p-1} & \boldsymbol{\gamma}_{23,p-1} \\ \boldsymbol{\gamma}_{31,p-1} & \boldsymbol{\gamma}_{32,p-1} & \boldsymbol{\gamma}_{33,p-1} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{1t-p+1} \\ \Delta \mathbf{x}_{2t-p+1} \\ \Delta \mathbf{x}_{3t-p+1} \end{bmatrix}$$

$$+ \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1' & \boldsymbol{\beta}_2' & \boldsymbol{\beta}_3' \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1t-p} \\ \mathbf{x}_{2t-p} \\ \mathbf{x}_{3t-p} \end{bmatrix} + \boldsymbol{\varepsilon}_t , \qquad (4.116b)$$

where $\boldsymbol{\gamma}_{n\,ij}$ has dimensions $N_n \times N_i$, $\boldsymbol{\alpha}_n$ is $N_n \times r$, and $\boldsymbol{\beta}_n'$ is $r \times N_n$, for n and i equal to 1, 2 or 3. The null hypothesis that $\mathbf{x}_{3t}$ does not Ganger cause $\mathbf{x}_{1t}$ can be set up either in terms of the levels *VAR* as:

$$H_0 : \; \boldsymbol{\phi}_{13,1} = \boldsymbol{\phi}_{13,2} = \cdots = \boldsymbol{\phi}_{13,p} = \mathbf{0}, \qquad (4.117a)$$

or in terms of the *VECM* as:

$$H_0^* : \; \boldsymbol{\gamma}_{13,1} = \boldsymbol{\gamma}_{13,2} = \cdots = \boldsymbol{\gamma}_{13,p} = \mathbf{0} \quad \text{and} \quad \boldsymbol{\alpha}_1 \boldsymbol{\beta}_3' = \mathbf{0}, \qquad (4.117b)$$

where the first part of the hypothesis relates to short-run noncausality, and the second part relates to long-run noncausality (or neutrality).

Toda and Phillips (1993b, 1994) have shown that a Wald test of H_0, based on the levels *VAR*, has an asymptotic chi-squared distribution if rank($\boldsymbol{\alpha}_3$) = N_3. They have also proposed 3 different Wald tests of H_0^*, based on the *VECM*, which asymptotically follow a chi-squared distribution, and which are applicable depending on which of the following set of assumptions holds: (1) rank($\boldsymbol{\alpha}_1$) = N_1 or rank($\boldsymbol{\beta}_3$) = N_3; (2) $\boldsymbol{\alpha}_1 = 0$; or (3) $\boldsymbol{\beta}_3 = 0$. The formulae to construct these statistics are given in Toda and Phillips (1994), and Granger non-causality tests with standard asymptotic distributions can be applied using some sequential procedure to identify whether the necessary requirements are consistent with the evidence.

4.4.3.2.7 Johansen's procedure and other optimal estimators of cointegrating relationships

Johansen's MLECM procedure and the single equation's optimal estimators discussed in section 4.3.3.1.4 offer different alternatives to estimate efficiently cointegrating relationships between *I*(1) variables. Keeping in mind the issues pertaining to the lack of identification of the structural cointegrating vectors when it is not unique --except if some *a priori* conditions have been imposed, it is important to aknowledge the relative performance of these different methods in finite samples or for innovations that are non-Gaussian or that have some general dynamics.

Stock and Watson (1993) have conducted some Monte Carlo experiments with bivariate models. They found that in finite samples of 100 observations, for some of their simulations, dynamic GLS and the fully-modified LS show biases, while Johansen's estimator is unbiased but its empirical distribution is wider. By comparison, dynamic OLS has the lowest root mean-squared error. Independently, Gonzalo's (1994) Monte Carlo study with a bivariate model suggests that Johansen's MLECM keeps desirable properties with nonnormal errors or slight inadequacies of the specified dynamics. Further work would be required to assess the finite sample behaviour of these various estimators with a largest number of variables. Pitarakis' (1993) study of cointegrating rank tests shows that inference about r tends to be distorted when the ratio of the number of variables to the sample size increases. This may be suggestive of problems in finite samples.

Before illustrating the use of Johansen's MLECM procedure with a specific example, it is helpful to summarize the steps involved in its application.

1. Examine the data to detect visually the possible presence of a trend.
2. Apply unit root tests on each variable.
3. Choose the lag length of the *VECM* system, i.e. p. Choose a representation that yields residuals for which we cannot reject the absence of serial correlation and their normality. The former can be tested with *LM*-tests for serial correlation, while the latter can be assessed with the Bera-Jarque test.
4. Maintaining the lag length p, apply cointegrating rank tests. In particular, use the *PGp*-statistic. You can also use the information criterion *ICPG*. Maintaining p and r, apply tests about the presence or not of a linear trend.
5. Examine graphs of the residual vectors associated with the r cointegrating vectors, i.e. $\beta_i'X_t$ for $i = 1, ..., r$, where β_i is the i^{th} column of β. Examine also graphs of the residual vectors associated with the r cointegrating vectors adjusted for short-run dynamics, i.e. $\beta_i'\mathbf{R}_{pt}$. These graphs should be consistent with the existence of long-run relationships in agreement with the rank r.

6. Perform any hypothesis testing of interest about the adjustment matrix α or the cointegrating space spanned by β.

4.4.3.3 Example:inflationary expectations and rationality II

Panels A and B of table 4.8 show respectively the levels and the first-differences of both the actual and the expected inflation series from 1951H2 to 1990H1. The evolution of these series does not suggest a linear trend in the *DGP* of the series. We therefore proceed with the specification of the *VECM* given by case 4 in table 4.7. Unit root tests on the series were previously conducted and reported in section 4.3.4. They provided evidence supporting first order integration in the levels of both series.

Table 4.9 shows some statistics for the normality and the i.i.d. assumptions for the residuals in model (4.75) (without linear trend) with $p = 6$. This specification was the most parsimonious one that did not suggest problems of serial correlation.

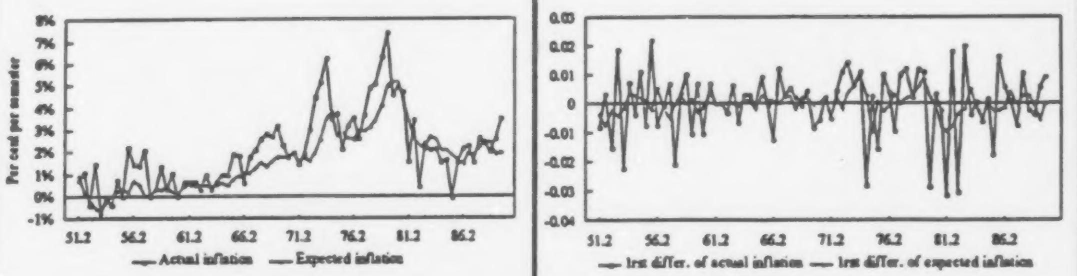
Table 4.10 shows different cointegating rank statistics discussed in section 4.4: *MV, TR, RATR, PG, PGp, ICPG.* All these test statistics reject the null hypothesis of no cointegration relation in favour of the alternative hypothesis of having at least one cointegrating relation at the 1% significance level. The null hypothesis of having at most one cointegrating relationship against the alternative of having two cointegrating relations is not rejected at conventional statistical levels.[70] The *ICPG* information criterion suggested by Pitarakis (1993) favours also a cointegrating rank equal to 1. We conclude therefore that there is one cointegrating vector whose basis is given by the eigenvector associated with the largest eigenvalue of matrix Π.

An examination of the residuals from the cointegrating relationship and of the residuals of the cointegrating relationship adjusted for short-run dynamics, as illustrated in panels C and D of table 4.8, is consistent with the tests' results. Both residuals series appear to be stationary. By contrast, as shown in panels E and F of table 4.8, the generated residuals associated with the second eigenvector of matrix Π do not appear to be stationary, whether or not they are adjusted for short-run dynamics. This second eigenvector does not lie in the cointegating space.

---

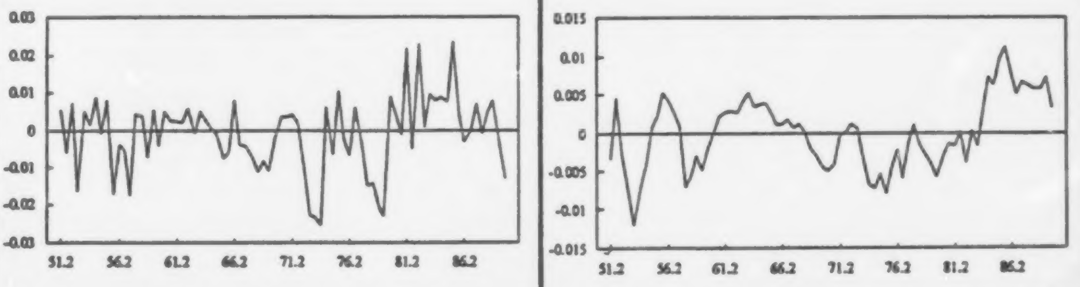
[70] Since the system under study has $N = 2$ variables, if the cointegrating rank had been 2, it would mean that the two series were stationary to begin with. Recall that for two integrated variables, there can be at most one unique cointegrating vector.

Table 4.8 - Actual and expected inflation data, cointegrating relationship, and non-cointegrating relationship. Sample 1951H2- 1990H1.



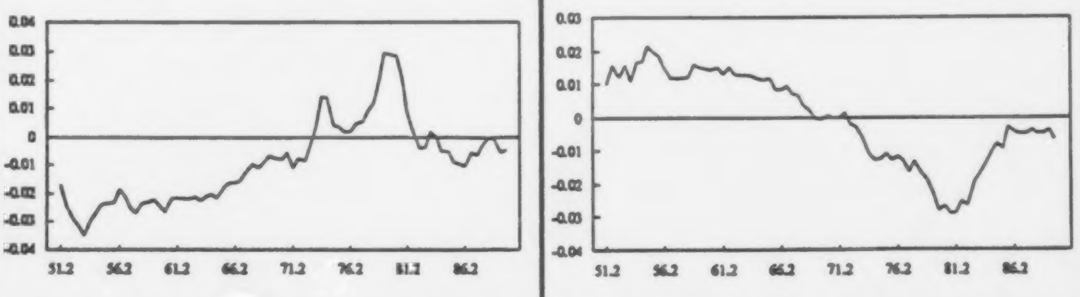
A) Actual and Livingston measure of expected inflation

B) First-difference of actual inflation and first-difference of expected inflation.



C) Unique cointegrating relationship (r=1) between the two series

D) Residuals of the cointegrating relationship (r=1) adjusted for short-run dynamics



E) Non-cointegrating relationship based on the eigenvector of the 2nd largest eigenvalue (r=2)

F) Residuals of the non-cointegrating relationship based on the eigenvector of the 2nd largest eigenvalue (r=2)

Table 4.9 - Statistics on the specification of a *VAR* model with intercept and lag lentgh of order 6: for actual and expected US inflation rates- 1951H2-1990H1.[1]

Dependent variable in the equation	S.E. of regression	LM tests fos serials correlation			Normality test (2)	RESET(3)	ARCH(2)	ARCH(4)	White test for heteros. (25)
		2	6	12					
PI^e_t	0.0030	4.28	10.03	17.03	1.12	0.355	1.26	3.56	37.73
	(p-value)	0.118	0.124	0.149	0.571	0.949	0.532	0.469	0.037
PI_t	0.0101	3.069	8.20	10.26	1.72	3.02	4.79	5.18	50.45
	(p-value)	0.217	0.224	0.593	0.423	0.388	0.091	0.270	0.001

[1] The White test for heteroskedasticity was conducted with the regressors and their squares.

Table 4.10 - Various cointegrating rank statistics for Johansen's ML procedure with intercept in the cointegrating vector: for actual and expected US inflation rates; 1951H2-1990H1.[1]

	MV	*TR*	*RATR*	*PG*	*PGp*	*ICPG*
H_0: $r = 0$	20.94*	23.69*	21.86*	22.37*	21.14*	9.97
H_0: $r \leq 1$	2.75	2.75	2.53	2.72	2.42	**-0.38**

[1] Rejection level: * at 1%; ** at 5%; *** at 10%.
Source for the critical values: Table A4.3
The alternative hypothesis for *MV* is H_1: $r = 1$ in the first row, and H_1: $r = 2$ in the second row.
For the other statistics, the alternative hypothesis is H_1: $r \geq 1$ in the first row, and H_1: $r = 2$ in the second row.
A bold figure indicates that an information criterion favours this hypothesis.

Table 4.11 - Estimation results from Johansen's ML procedure for actual and expected US inflation rates; 1951H2-1990H1

Eigenvalues	$\tau_1 = 0.23545$,	$\tau_2 = 0.0346$,	$\tau_3 = 0$					
Variables	Normalized eigenvectors associated with		Normalized estimated adjustment matrix α for $r = 1$	Estimated Long Run Matrix Π for $r = 1$			Estimated adjustment matrix α for $r = 1$ from marginal models	*p*-value for Wald test for signif. of α_i $\chi^2(1)$
	τ_1	τ_2		PI^e_t	PI_t	I		
PI^e_t	1.0000	1.0000	-0.13082	-0.13082	0.13466	-0.00528	-0.12046	0.093
PI_t	-1.0293	0.0966	0.64274	0.62340	-0.64168	0.00252	0.65496	0.008
I	0.00404	-0.0273						

For $r = 1$, and $p = 6$, we have conducted a test of the null hypothesis that there is no time trend in the DGP of the model. The likelihood ratio given by equation (4.103) was computed as:

$$LR_{\theta_0} = -78 \sum_{i=2}^{2} \log\left[\frac{1-0.034587}{1-0.023403}\right] = 0.8984 \sim \chi^2(1), \qquad (4.118)$$

whose p-value is 0.3432. Hence, we do not reject the absence of linear time trend in our specification at the 34% significance level. The adopted *VECM* specification is therefore given by:

$$\Delta \mathbf{X}_t = \sum_{i=1}^{5} \Gamma_i \, \Delta \mathbf{X}_{t-i} + \alpha \, \beta' \, \mathbf{X}_{t-6} + \varepsilon_t, \qquad (4.119a)$$

or

$$\begin{bmatrix} \Delta PI^e_t \\ \Delta PI_t \end{bmatrix} = \sum_{i=1}^{5} \Gamma_i \begin{bmatrix} \Delta PI^e_{t-i} \\ \Delta PI_{t-i} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 & \delta_0 \end{bmatrix} \begin{bmatrix} PI^e_{t-6} \\ PI_{t-6} \\ 1 \end{bmatrix} + \varepsilon_t . \qquad (4.119b)$$

Table 4.11 gives estimates of the normalized cointegrating vector (up to a scaling factor), of the second eigenvector, of the normalized adjustment matrix α, and of the long-run matrix Π. It also provides estimates of the elements of the adjustment matrix α, that were obtained from OLS estimation of the marginal models. Because each equation was estimated separately, taking as given the estimates of β obtained by FIML estimation of the system, these estimates of the marginal models may be less precise even though they are consistent. The point estimates obtained by both the system estimation and the estimation of the marginal models are fairly similar. We then conducted tests of weak exogeneity of the variables for the parameters of interest in each partial model. These tests revealed that the error-correction term appears to be significant at the 10% level in the PI^e_t equation, and at the 1% significance level in the PI_t equation. Hence, they do not support the weak exogeneity of any of the series regarding the long-run coefficients in either equation of the system. This would therefore favour the use of a system procedure to apply cointegration analysis on PI^e_t and PI_t.

We next performed tests on the elements of the vector generating the cointegrating space. First, we tested the linear homogenous restriction that $\beta_1 + \beta_2 = 0$. This can be formulated as:

$$H_0 : \beta = \mathbf{H}_{(2+1)\times 2}\,\varphi_{2\times 1} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varphi_{11} \\ \varphi_{01} \end{bmatrix} = \begin{bmatrix} \varphi_{11} \\ -\varphi_{11} \\ \varphi_{10} \end{bmatrix} . \qquad (4.120)$$

Because there is an intercept in the cointegrating vector, matrix **H** must be of dimension $(N+1) \times s$. The likelihood ratio statistic defined by equation (4.114) can be calculated and follows a χ^2-distribution with (1 (2 - 1)) = 1 degree of freedom. The statistic LR_β is equal to 0.1338, which does not reject H_0 even at a 71% significance level.

We also fixed the cointegrating vector to be equal to [1, -1, 0]. The *LR* test of this null hypothesis was equal to 6.50, which follows a χ^2-distribution with $(r, (N+1-r))$ = (1 (2+1 - 1)) = 2 degrees of freedom. The *p*-value of this statistic is 0.039, thus rejecting the null hypothesis at a 5% significance level.

Finally, we fixed the elements of the cointegrating vector to be equal to [1, -1, -0.0003]. The relevant *LR* statistic was then equal to 5.83, with a probability value of 0.054 according to a χ^2-distribution with 2 degrees of freedom. This suggests some lower bound on the value of the intercept in the cointegrating relation. Such lower bound is consistent with the value reported in tables 4.5 and 4.6. It represents an average gap between actual and expected inflation of only about six tenths of 1% yearly.

On this statistical ground, such a small intercept and the non-rejection of a cointegrating relation between PI^e_t and PI_t that is proportional to (1,-1) is consistent with the rational expectations hypothesis. In this particular example with two variables, the cointegration analysis performed on an individual equation (see section 4.3.4) and that performed on a system (in this section) gave similar results, even though it was found that the integrated series were endogenous. This will not necessarily always be the case, so that some caution is warranted.

4.5 Summary.

As macroeconomics seeks to explain individual movements and comovements in macroeconomic variables over time, dynamic models constitute a cornerstone in modern macroeconometrics. The recent econometric literature has developped in many ways to deepen our understanding of dynamic models. They are useful both as mere statistical representations of the data and as a framework that anchors the dynamics in a manner consistent with long-run equilibrium relationships suggested by economic theory.

In this chapter, we have shown the existence of many equivalent reparameterisations of dynamic models, which can prove valuable to emphasize different sets of parameters, according to one's interest. In some instances, individual dynamic equations are very convenient and instructive. In others, vector autoregressive representation (*VAR*) and vector error-correction model (*VECM*) are atheoretical macroeconomic tools in a multivariate context, that are helpful to characterize the data or to construct optimal forecasts.

In any case, it remains important to know one's data when dealing with econometric modelling. In particular, one has to verify whether some or all the series under consideration are integrated, and whether there may be cointegrating relationships between the non-stationary variables (regressors or regressand). Indeed, for a specification to be meaningful, or not to reflect some spurious relationships, it is required that balanced regressions be estimated. Furthermore, complications about inference on some parameters may arise from the presence of cointegrating regressors. In this context, one must verify that some econometric results can justify the reference to standard asymptotic distribution. If this is not the case, proper inference on coefficients of integrated regressors will require Monte Carlo simulation to assess the empirical distribution of the test statistics.

The procedure below summarizes how to deal systematically with issues of cointegration:

1. Perform unit root tests on each variable of the system (see section 3.3.3).
2. Apply Johansen's procedure, with Pitarakis' (1993) set of tools to determine the number of cointegrating relationships, and to test for failure of weak exogeneity.

3. If you find or if you suspect problems regarding the weak exogeneity of the regressors, estimate efficiently with full information MLE. The other option, if you decide working in the context of an individual equation is to use an efficient modified estimator (see section 4.3.3.1.4).

4. If you do not suspect endogeneity, you can work with the relevant static cointegrating regressions, and you can apply residual based tests for cointegration, especially ADF and Phillips-Perron $Z(\rho)$ tests. Be aware however that static cointegrating regressions may suffer from finite sample biases. As an alternative, if you are interested in the long-run multipliers, consider estimating an ECM representation or its Bewley's transformation. In any case, remember that for N greater than 2, the cointegrating vector may not be unique (which may have been checked with Johansen's procedure) so that you are only estimating vectors that are contained in the cointegrating space.

4.6 *Appendix Tables.*

Table A4.1.- Asymptotic critical values of no cointegration tests based on response surfaces[1]

Test statistic[2]	Size of the test			
	0.01	0.025	0.05	0.10
$N = 2$				
$t_{\rho c}$	-3.90	-3.59	-3.34	-3.04
$t_{\rho ct}$	-4.32	-4.03	-3.78	-3.50
$t_{\rho ctt}$	-4.69	-4.40	-4.15	-3.87
$T(\hat{\rho}_c - 1)$	-28.3	-23.9	-20.6	-17.1
$T(\hat{\rho}_{ct} - 1)$	-35.8	-31.1	-27.3	-23.4
$T(\hat{\rho}_{ctt} - 1)$	-42.6	-37.5	-33.4	-29.1
$N = 3$				
$t_{\rho c}$	-4.29	-4.00	-3.74	-3.45
$t_{\rho ct}$	-4.66	-4.37	-4.12	-3.84
$t_{\rho ctt}$	-4.99	-4.70	-4.45	-4.17
$T(\hat{\rho}_c - 1)$	-35.2	-30.4	-26.7	-22.7
$T(\hat{\rho}_{ct} - 1)$	-42.0	-36.9	-32.8	-28.5
$T(\hat{\rho}_{ctt} - 1)$	-48.5	-43.0	-38.7	-34.0
$N = 4$				
$t_{\rho c}$	-4.64	-4.35	-4.10	-3.81
$t_{\rho ct}$	-4.97	-4.68	-4.43	-4.15
$t_{\rho ctt}$	-5.27	-4.98	-4.73	-4.45
$T(\hat{\rho}_c - 1)$	-41.6	-36.5	-32.4	-28.1
$T(\hat{\rho}_{ct} - 1)$	-48.1	-42.6	-38.2	-33.5
$T(\hat{\rho}_{ctt} - 1)$	-54.3	-48.5	-43.9	-38.9
$N = 5$				
$t_{\rho c}$	-4.96	-4.66	-4.42	-4.13
$t_{\rho ct}$	-5.25	-4.96	-4.72	-4.43
$t_{\rho ctt}$	-5.53	-5.24	-4.99	-4.72
$T(\hat{\rho}_c - 1)$	-47.8	-42.3	-38.0	-33.3
$T(\hat{\rho}_{ct} - 1)$	-54.0	-48.2	-43.5	-38.5
$T(\hat{\rho}_{ctt} - 1)$	-60.0	-53.9	-49.0	-43.7

Table A4.1.- Asymptotic critical values of no cointegration tests based on response surfaces[1]

Test statistic[2]	Size of the test			
	0.01	0.025	0.05	0.10
		$N = 6$		
$t_{\rho c}$	-5.25	-4.96	-4.71	-4.42
$t_{\rho ct}$	-5.52	-5.23	-4.98	-4.70
$t_{\rho ctt}$	-5.77	-5.49	-5.24	-4.96
$T(\hat{\rho}_c - 1)$	-53.8	-48.0	-43.4	-38.4
$T(\hat{\rho}_{ct} - 1)$	-59.7	-53.7	-48.8	-43.5
$T(\hat{\rho}_{ctt} - 1)$	-65.5	-59.2	-54.1	-48.6

[1] Source: Davidson and MacKinnon (1993).

[2] The indices *c*, *t*, and *tt* respectively mean that the cointegrating regression includes a constant, both a constant and a deterministic time trend, and all of a constant, t and t^2.

Table A4.2- Quantile of the asymptotic distribution of the cointegrating rank statistics (Case 3)[1]

DGP and statistical model:

$$\Delta X_t = \alpha \delta_0 + \alpha_\perp \theta_0 + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-p+1} + \Pi X_{t-p} + B Z_t + \varepsilon_t$$

with $\delta_0 \neq 0$, $\theta_0 \neq 0$, and $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$

$N-r$	90%	95%	97.5%	99%
		MV		
1	2.69	3.76	4.95	6.65
2	12.07	14.07	16.05	18.63
3	18.60	20.97	23.09	25.52
4	24.73	27.07	28.98	32.24
5	30.90	33.46	35.71	38.77
6	36.76	39.37	41.86	45.10
7	42.32	45.28	47.96	51.57
8	48.33	51.42	54.29	57.69
9	53.98	57.12	59.33	62.80
10	59.62	62.81	65.44	69.09
11	65.38	68.83	72.11	75.95
		TR		
1	2.69	3.76	4.95	6.65
2	13.33	15.41	17.52	20.04
3	26.79	29.68	32.56	35.65
4	43.95	47.21	50.35	54.46
5	64.84	68.52	71.80	76.07
6	89.48	94.15	98.33	103.18
7	118.50	124.24	128.45	133.57
8	150.53	156.00	161.32	168.36
9	186.39	192.89	198.82	204.95
10	225.85	233.13	239.46	247.18
11	269.96	277.71	284.87	293.44

[1] Source: Osterwald-Lenum (1992), table 1.

Table A4.3- Quantile of the asymptotic distribution of the cointegrating rank statistics (Case 4)[1]

DGP and statistical model:

$$\Delta X_t = \alpha \delta_0 + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-p+1} + \Pi X_{t-p} + B Z_t + \varepsilon_t$$

with $\delta_0 \neq 0$, $\theta_0 = 0$, and $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$

N - r	90%	95%	97.5%	99%
		MV		
1	7.52	9.24	10.80	12.97
2	13.75	15.67	17.63	20.20
3	19.77	22.00	24.07	26.81
4	25.56	28.14	30.32	33.24
5	31.66	34.40	36.90	39.79
6	37.45	40.30	43.22	46.82
7	43.25	46.45	48.99	51.91
8	48.91	52.00	54.71	57.95
9	54.35	57.42	60.50	63.71
10	60.25	63.57	66.24	69.94
11	66.02	69.74	72.64	76.63
		TR		
1	7.52	9.24	10.80	12.97
2	17.85	19.96	22.05	24.60
3	32.00	34.91	37.61	41.07
4	49.65	53.12	56.06	60.16
5	71.86	76.07	80.06	84.45
6	97.18	102.14	106.74	111.01
7	126.58	131.70	136.49	143.09
8	159.48	165.58	171.28	177.20
9	196.37	202.92	208.81	215.74
10	236.54	244.15	251.20	257.68
11	282.45	291.40	298.31	307.64

[1] Source: Osterwald-Lenum (1992), table 1*.

Table A4.4- Quantile of the asymptotic distribution of the cointegrating rank statistics (Case 4a)[1]

DGP

$\Delta X_t = \alpha\delta_0 + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-p+1} + \Pi X_{t-p} + B Z_t + \varepsilon_t$

with $\delta_0 \neq 0$, $\theta_0 = 0$, and $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$

Statistical model

$\Delta X_t = \alpha\delta_0 + \alpha_\perp \theta_0 + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-p+1} + \Pi X_{t-p} + B Z_t + \varepsilon_t$

allowing both $\delta_0 \neq 0$, $\theta_0 \neq 0$

N - r	90%	95%	97.5%	99%
		MV		
1	6.50	8.18	9.72	11.65
2	12.91	14.90	17.07	19.19
3	18.90	21.07	22.89	25.75
4	24.78	27.14	29.16	32.14
5	30.84	33.32	35.80	38.78
6	36.35	39.43	41.86	44.59
7	42.06	44.91	47.59	51.30
8	48.43	51.07	53.85	57.07
9	54.01	57.00	59.80	63.37
10	59.19	62.42	64.98	68.61
11	65.07	68.27	70.69	74.36
		TR		
1	6.50	8.18	9.72	11.65
2	15.66	17.95	20.08	23.52
3	28.71	31.52	34.48	37.22
4	45.23	48.28	51.54	55.43
5	66.49	70.60	74.04	78.87
6	90.39	95.18	99.32	104.20
7	118.99	124.25	129.75	136.06
8	151.38	157.11	162.75	168.92
9	186.54	192.84	198.06	204.79
10	226.34	232.49	238.26	246.27
11	269.53	277.39	283.84	292.65

[1] Source: Osterwald-Lenum (1992), table 1.1*.

5. CONCLUSIONS.

At the beginning of this guide, it was argued that applied macroeconometrics has been set on a new course because of advances in econometric theory and the development of ever more powerful computers that make use of new techniques more readily applicable and programmable. While a thorough understanding of the fundamental statistical theory underlying these new techniques may be too involved for the needs and interests of most practitioners, a definite goal of this guide has been to make accessible a heuristic and step-by-step presentation of modern macroeconometric techniques. The application and the interpretation of the new techniques can be generally understood intuitively in terms of introductory econometrics concepts, in a fairly unified approach to applied econometrics.

The evolution of the econometric approach before and after the 1980's can be illustrated schematically in Figure 5.1 and Figure 5.2.[1] Figure 5.1 represents a rather one way methodology that uses economic theory to specify at the outset a tight and parsimonious model. It is subsequently estimated, and it is hastily used to test economic hypotheses of interest. Finally, the estimated model is used for predictions and policy analysis. Basically, such an approach is deficient in many ways. It fails to understand the fundamental nature and properties of the data at hands. It does not seek to specify a general model that may embed more restricted models. It does not assess the adequacy of the formulated model and its specification. It lacks feedbacks from econometric testing to economic theory, to the empirical specification and to the data.

By contrast, Figure 5.2 proposes a modified schematic description that is in line with the econometric approach exposed throughout this guide. This approach is derived from both the advances in the modern econometric literature, and from our experience with statistical analysis of macroeconomic data. Following Hendry (1979), economic theory is employed to formulate a fairly general dynamic empirical model. Then, the data are examined for the presence of unit roots. If the evidence suggests that the variables are stationary, this may influence the formulation of the specification to be estimated. If some or all the variables are integrated (say of order one), we then need to test for possible cointegrating, or long-run stationary, relations between these variables. Such cointegrating relationships must be taken into account in one way or another in the econometric specification to be considered. The absence of cointegrating relationships would lead either to transforming the variables in a stationary form before estimating an appropriate specification, or to going back to think harder about economic theory.

---

[1] These figures were modified and adapted by the author from Maddala (1992).

Once a general econometric specification is set up in a way compatible with the nature of the data, the model can be estimated. The next stage involves the application of a battery of misspecification tests to assess the statistical adequacy of the empirical model. It is at this stage that the checklist presented in chapter 2 section 4 must play a role. If the reasearcher can comfortably argue that the model is statistically sound, he or she can proceed with the economic analysis of the empirical model and, if appropriate, use it for predictions and policy purposes.

This modern approach remains certainly demanding and offers no panacea for careful and methodical econometric work. It is not perfect either, and it will always require at times borderline decisions about which way to take to develop a better model.[2] Because of the criteria it employs, this approach seeks however to cut down on arbitrariness by adding more rigour in the empirical quest. In this sense, it forces more a hands-on-the-table econometric approach. The application of the new techniques has certainly brought significant progress in the modeling and the understanding of macroeconomic variables. Their understanding and their use are now a must.

---

[2] Replying to Hendry and Ericsson (1991), Friedman and Schwartz (1991) build up a good case for knowing one's data, being well aware of the errors-in-variables problem, and ultimately assessing how the model successfully fits data out of the estimation sample.

FIGURE 5.1 An out-of-date and faulty econometric approach.

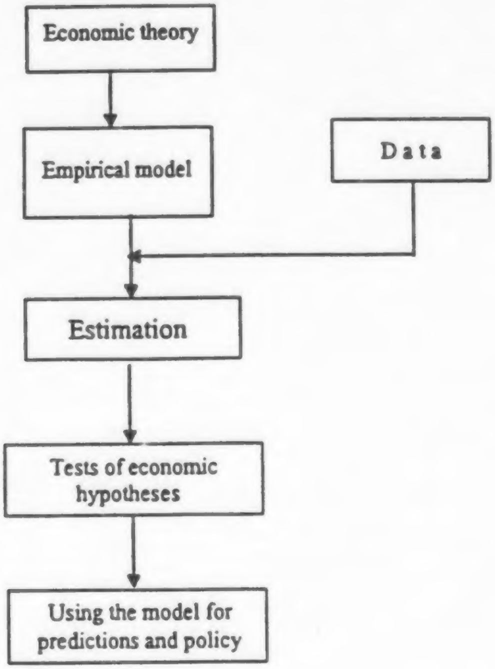
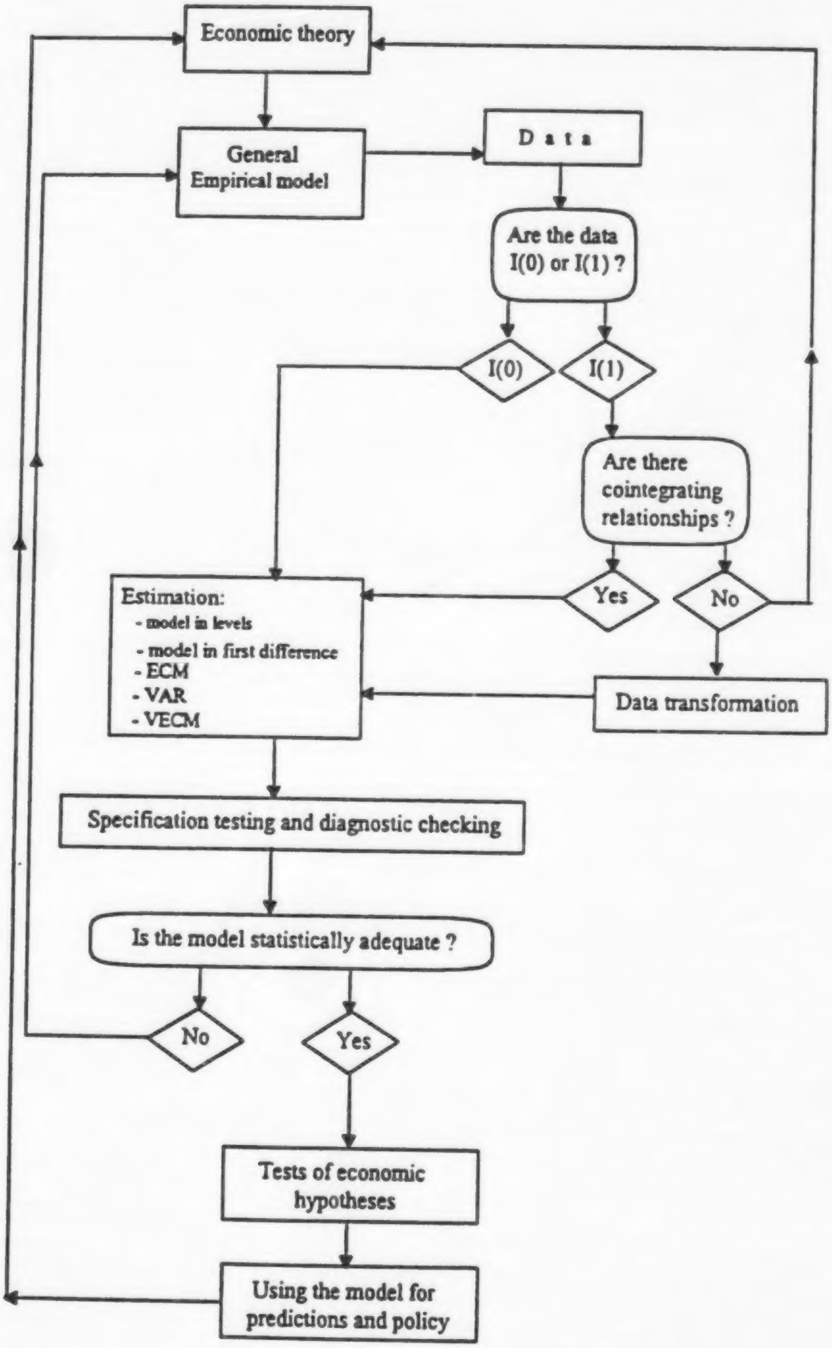

Economic theory
Empirical model
Data
Estimation
Tests of economic hypotheses
Using the model for predictions and policy

FIGURE 5.2 A modern approach to economet


Economic theory
General Empirical model
Data
Are the data I(0) or I(1) ?
I(0)
I(1)
Are there cointegrating relationships ?
Yes
No
Estimation:
- model in levels
- model in first difference
- ECM
- VAR
- VECM
Data transformation
Specification testing and diagnostic checking
Is the model statistically adequate ?
No
Yes
Tests of economic hypotheses
Using the model for predictions and policy

References

AKAIKE, H. (1970), "Statistical Predictor Identification," Annals of the Institute of Statistics and Mathematics 22, pp. 203-217.

AKAIKE, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in Petrov-Csaki ed., Second International Symposium on Information Theory, Budapest, Akademia Kiado.

AMBLER, S. and A. PAQUET (1992), "Cointegration and the Demand for M2 and M2+ in Canada," in D. Longworth ed., Monetary Seminar 90, Bank of Canada, pp. 125-168.

AMBLER, S. and L. PHANEUF (1989), "Aggregate Disturbances, Dynamic Propagation and the Persistence of Output and Unemployment Fluctuations," Cahier de recherche no. 3, Centre de recherche sur les politiques économiques, Université du Québec à Montréal.

AMEMIYA, T. (1980), "Selection of Regressors," International Economic Review 21, pp. 331-354.

ANDREWS, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Econometrica 59, pp. 817-858.

ANDREWS, D.W.K. and J.C. MONAHAN (1991), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," Econometrica 60, pp. 953-966.

BANERJEE, A., R.L. LUMSDAINE, and J.H. STOCK (1992), "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," Journal of Business and Economic Statistics 10, pp. 271-287.

BANERJEE, A., J. DOLADO, J.W. GALBRAITH, and D.F. HENDRY (1993), Co-integration, Error-Correction, and the Analysis of Non-stationary Data, Oxford, Oxford University Press.

BANERJEE, A., J.J. DOLADO, D.F. HENDRY, and G.W. SMITH (1986), "Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence," Oxford Bulletin of Economics and Statistics 48, pp. 253-277.

BARTLETT, M.S. (1946), "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series," Journal of the Royal Statistical Association B8, pp. 27-41.

BERA, A.K. and C.M. JARQUE (1981), "An Efficient Large-Sample Test for Normality of Observations and Regression Residuals," Australian National University working papers in econometrics No. 40, Canberra.

BERNANKE, B. (1986), "Alternative Explanations of the Money-Income Correlation," Carnegie-Rochester Conference Series on Public Policy 25, pp. 49-99.

BERNDT, E.R. and N.E. SAVIN (1977), "Conflict among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," Econometrica 45, pp. 1263-1278.

BLANCHARD, O.J. and D. QUAH (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances," American Economic Review 79, pp. 655-673.

BLANCHARD, O.J. and M.W. WATSON (1986), "Are Business Cycles All Alike?," in R.J. Gordon ed., The American Business Cycle: Continuity and Change, Chicago, University of Chicago Press, pp. 123-179.

BOLLERSLEV, T. (1986), "Generalized Autoregressive Heteroskedasticity," Journal of Econometrics 31, pp. 307-327.

BOOTHE P. (1992), "Commentary on: A BVAR Model for Forecasting and Analysis of Canadian Monetary Policy by Racette and Raynauld," in D. Longworth ed., Monetary Seminar 90, Bank of Canada, pp. 387-392.

BOX, G.E.P. and G.M. JENKINS (1970), Time Series Analysis, Forecasting and Control, San Francisco: Holden Day.

BRAUN, P.A. and S. MITTNIK (1993), "Misspecifications in Vector Autoregressions and their Effects on Impulse Responses and Variance Decompositions," *Journal of Econometrics* 59, pp. 319-341.

BREUSCH, T.S. (1978), "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers* 17, pp. 334-355.

BREUSCH, T.S. and L.G. GODFREY (1981), "A Review of Recent Work on Testing for Autocorrelation in Dynamic Simultaneous Models," in D. Currie, R. Nobay, and D. Peel eds., *Macroeconomic Analysis: Essays in Macroeconomics and Econometrics*, London, Crook Helm.

CAMPBELL, J.Y. and P. PERRON (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," in S. Fisher ed., *NBER Macroeconomics Annual 1991*, Cambridge, MIT Press, pp. 141-201.

CHOI, I. (1994), "Spurious Regressions and Residual-Based Tests for Cointegration when Regressors Are Cointegrated," *Journal of Econometrics* 60, pp. 313-320.

CHOI, I. and P.C.B. PHILLIPS (1993), "Testing for a Unit Root by Frequency Domain Regression," *Journal of Econometrics* 59, pp. 263-286.

CHRISTIANO, L.J. (1992), "Searching for a Break in GNP," *Journal of Business and Economic Statistics* 10, pp. 237-250.

CHRISTIANO, L.J. and M. EICHENBAUM (1990), "Unit Roots in Real GNP : Do we Know, and do we Care?," *Carnegie-Rochester Conference Series on Public Policy* 32, pp.7-62.

CHRISTIANO, L.J. and M. EICHENBAUM (1992), "Identification and the Liquidity Effect of a Monetary Policy Shock," in A. Cukierman, Z. Hercowitz, and L. Leiderman eds., *Political Economy, Growth, and Business Cycles*, Cambridge, MIT Press, pp. 335-370.

CHRISTIANO, L.J., M. EICHENBAUM and C. Evans (1994), "The Effects of Monetary Policy Shocks: Some Evidence from the Flow of Funds," National Bureau of Economic Research Working Paper No. 4699, Cambridge.

COCHRANE, J.H. (1988), "How Big is the Random Walk in GNP?" *Journal of Political Economy* 96, pp. 893-920.

COOLEY, T.A. and S.F. LEROY (1985), "Atheoretical Macroeconomics: A Critique," *Journal of Monetary Economics* 16, pp. 283-308.

CRAVEN, P. and G. WAHBA (1979), " Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numerische Mathematik* 31, pp. 377-403.

CUKIERMAN, A. and A.H. MELTZER (1982), "What do Tests of Market Efficiency Show in the Presence of the Permanent-Transitory Confusion," mimeo., Tel-Aviv University and Carnegie-Mellon University, Pittsburgh, PA.

DAVIDSON, R. and J.G. MACKINNON (1993), *Estimation and Inference in Econometrics*, Oxford, Oxford University Press.

DeJONG, D.N., J.C. NANKERVIS, N.E. SAVIN, and C.H. WHITEMAN (1992), "Integration versus Trend Stationarity in Time Series," *Econometrica* 60, pp. 423-433.

DICKEY, D.A. (1976), "Estimation and Hypothesis Testing for Nonstationary Time Series," Ph.D. dissertation, Iowa University.

DICKEY, D.A., W.R. BELL, and R.B. MILLER (1986), "Unit Roots in Time Series Models: Tests and Applications," *American Statistician* 40, pp. 12-26.

DICKEY, D.A. and W.A. FULLER (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, pp. 427-431.

DICKEY, D.A. and W.A. FULLER (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica* 49, pp. 1057-1072.

DHRYMES, P.J. (1974), *Econometrics: Statistical Foundations and Applications*, New York, Springer-Verlag.

DOAN, T., R.B. LITTERMAN, and C. SIMS (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Review* 3, pp. 1-100.

DURBIN J. (1960), "The Fitting of Time Series Model," *Review of Inst. and International Statistics* 28, pp. 233-244.

ENGLE, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50, pp. 987-1007.

ENGLE, R.F. and S. BROWN (1985), "Model Selection for Forecasting," *Journal of Computation in Statistics*.

ENGLE, R.F. and C.W.J. GRANGER (1987), "Co-integration and Error Correction : Representation, Estimation, and Testing," *Econometrica* 55, pp. 251-276.

ENGLE, R.F., D.F. HENDRY, and J.F. RICHARD (1983), "Exogeneity," *Econometrica* 51, pp. 277-304.

EVANS, G.B.A. and N.E. SAVIN (1981), "Testing for Unit Roots : I," *Econometrica* 49, pp. 753-779.

FRIEDMAN, M. and A.J. SCHWARTZ (1991), "Alternative Approaches to Analyzing Economic Data," *American Economic Review* 81, pp. 39-85.

FULLER, W.A. (1976), *Introduction to Statistical Time Series*, New York: John Wiley & Sons.

GEWEKE, J. (1984), "Inference and Causality in Economic Time Series," in Z. Griliches and M.D. Intriligator eds., *Handbook of Econometrics* II, pp. 1101-1144.

GHYSELS, E. (1990), "On the Economics and Econometrics of Seasonality," mimeo., Université de Montréal.

GHYSELS, E. and E. KARANGWA (1988), "Nominal versus Real Seasonal Adjustments," Discussion Paper No. 2188, C.R.D.E., Université de Montréal.

GHYSELS, E. and P. PERRON (1993), "The Effects of Seasonal Adjustment Filters on Tests for a Unit Root," *Journal of Econometrics* 55, pp. 57-98.

GIANNINI, C. (1992), *Topics in Structural VAR Econometrics*, New-York, Springer-Verlag.

GODFREY, L.G. (1978), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables," *Econometrica* 46, pp. 1293-1302.

GODFREY, L.G. (1985), "On the Use of Misspecification Checks and Tests of Non-nested Hypotheses in Empirical Econometrics," *Economic Journal*, Supplement, pp. 69-81.

GODFREY, L.G. (1988), *Misspecification Tests in Econometrics : The Lagrange Multiplier Principle and Other Approaches*, Cambridge, Cambridge University Press.

GONZALO, J. (1994), "Five Alternative Methods of Estimating Long-run Equilibrium Relationships," *Journal of Econometrics* 60, pp. 203-233.

GRANGER, C.W.J. and P. NEWBOLD (1986), *Forecasting Economic Time Series*, 2nd edition, Academic Press.

GRANGER, C.W.J. and P. NEWBOLD (1974), "Spurious Regressions in Econometrics," *Journal of Econometrics* 2, pp. 111-120.

GREENE, W.H. (1993), *Econometric Analysis*, 2nd edition, New York, Macmillan.

GREGORY, A.W. (1991), "Testing for Cointegration in Linear Quadratic Models," Working Paper #811, Queen's University.

HALL, A. (1989), "Testing for a Unit Root in the Presence of Moving Average Errors," Biometrika 76, pp. 49-56.

HALL, A. (1992), "Testing for a Unit Root in Time Series Using Instrumental Variable Estimators with Pretest Data Based Model Selection," Journal of Econometrics 54, pp. 223-250.

HALL, R.E. (1978), "Stochastic Implications of the Life-Cycle Permanent Income Hypothesis," Journal of Political Economy 86, pp. 971-987.

HALL, S.G. (1989), "Maximum Likelihood Estimation of Cointegration Vectors: An Example of the Johansen Procedure," Oxford Bulletin of Economics and Statistics 51, pp. 213-218.

HAMILTON, J.D. (1994), Time Series Analysis, Princeton, Princeton University Press.

HANDA, J. and B.K. MA (1989), "Four Tests for the Random Walk Hypothesis: Power versus Robustness," Economics Letters 29, pp. 141-145.

HANNAN, E.J. and B. QUINN (1979), "The Determination of the Order of an Autoregression," Journal of the Royal Statistical Society B41, pp. 190-195.

HANNAN, E.J. and J. RISSANEN (1982), "Recursive Estimation of Mixed Autoregressive-Moving Average Order," Biometrika 69, 81-94. Correction (1983),70, p. 303.

HANSEN, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica 50, pp. 1029-1054.

HARVEY, A.C. (1981), Time Series Models, London: Philip Allan Publisher.

HARVEY, A.C. (1990), The Econometric Analysis of Time Series, 2nd edition, M.I.T. Press.

HAUSMAN, J.A. (1978), "Specification Tests in Econometrics," Econometrica 46, pp. 1251-1271.

HENDRY, D.F. (1979), "Predictive Failure and Econometric Modelling in Macroeconomics: The Transaction Demand for Money," in Paul Ormerod ed., Economic Modelling, ch. 9, London: Heinenmann, pp. 217-242.

HENDRY, D.F. (1980), "Econometrics - Alchemy or Science ?," Economica 47, pp. 387-406.

HENDRY, D.F. (1986), "Econometric Modelling with Cointegrated Variable : An Overview, " Oxford Bulletin of Economics and Statistics 48, pp. 210-212

HENDRY, D.F. and N.R. ERICSSON (1991), "An Econometric Analysis of U.K. Money Demand in *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna J. Schwartz," American Economic Review 81, pp. 8-38.

HODGSON, D. (1992), "Testing Time Series for Cointegration and the Estimation of Cointegrating Relationships with Emphasis on Johansen's Approach," mimeo., Department of Finance, Economic Analysis and Forecasting Division, Ottawa.

HOSTLAND, D. (1992), "Some Notes on the Use of Unit Root Testing Procedures in Applied Research," mimeo., Bank of Canada.

JARQUE, C.M. and C.R. McKENZIE (1983), "Testing for Multivariate Normality in Simultaneous Equation Models," Australian National University Working Papers in Economics and Econometrics No. 082.

JOHANSEN, S. (1988), "Statistical Analysis of Cointegration Vectors," Journal of Economic Dynamics and Control, pp. 231-254.

JOHANSEN, S. (1991a), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," Econometrica 59, pp. 1551-1580.

JOHANSEN, S. (1991b), "A Statistical Analysis of Cointegration for I(2) Variables," mimeo., Institute of Mathematical Statistics, University of Copenhagen.

JOHANSEN, S. (1992a), "Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data," Journal of Policy Modeling 14, pp. 313-334.

JOHANSEN, S. (1992b), "A Representation of Vector Autoregressive Processes Integrated of Order 2," Econometric Theory 8, pp. 188-202.

JOHANSEN, S. and K. JUSELIUS (1990), "Maximum Likelihood Estimation and Inference on Cointegration–with Applications to the Demand for Money," Oxford Bulletin of Economics and Statistics 52, pp. 169-210.

JOHANSEN, S. and K. JUSELIUS (1992), "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK,"Journal of Econometrics 53, pp. 211-244.

KAHN, J.H. and M. OGAKI (1990), "A Chi-square Test for a Unit Root," Economics Letters 34, pp. 37-42.

KEATING, J.W. (1990), "Identifying VAR Models under Rational Expectations," Journal of Monetary Economics 25, pp. 453-476.

KEATING, J.W. (1992), "Structural Approaches to Vector Autoregressions," Federal Reserve Bank of St-Louis Review 74, pp. 37-57.

KENNEDY, P. (1992), A Guide to Econometrics, 2nd edition, Cambridge, M.I.T. press.

KIM, I.M. (1990), "Structural Change and Unit Roots," Ph.D. dissertation, University of Florida.

KING, R.G., C.I. PLOSSER, J. STOCK and M.W. WATSON (1991), "Stochastic Trends and Economic Fluctuations," American Economic Review 81, pp. 819-840.

KOENKER, R. (1981), "A Note on Studentizing a Test for Heteroskedasticity," Journal of Econometrics 17, pp. 107-112.

LEAMER, E.E. (1983), "Let's Take the Con out of Econometrics," American Economic Review 73, pp.31-43.

LITTERMAN, R.B. (1980), "A Bayesian Procedure for Forecasting using Vector Autoregression," M.I.T. Department of Economics Working Paper.

LITTERMAN, R.B. (1986), "Forecasting with Bayesian Vector Autoregressions," Journal of Business and Economic Statistics 4, pp. 25-37.

LO, A.W., and W.K. NEWEY (1985), " A Large Sample Chow Test for the Linear Simultaneous Equation," Economics Letters 18, pp. 351-353.

MACKINNON, J.G. (1991), "Critical values for Cointegration Tests," Ch. 13 in R.F. Engle and C.W.J. Granger eds., Long-Run Economic Relationships: Readings in Cointegration, Oxford, Oxford University Press.

MACKINNON, J.G. and H. WHITE (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," Journal of Econometrics 29, pp. 305-325.

MADDALA, G.S. (1992), Introduction to Econometrics, New York, Macmillan.

MAEKAWA, K. (1994), "Prewhitened Unit Root Test," Economics Letters 45, pp. 145-153.

MALLOWS, C.L. (1973), "Some Comments on C_p," Technometrics 15, pp. 661-676.

MARDIA, K.V. (1980), "Tests for Univariate and Multivariate Normality," in P.R. Krishnaiah ed., Handbook of Statistics, Volume 1, Amsterdam, North Holland.

McALEER, M., A.R. PAGAN, and P.A. VOLKER (1985), "What Will Take the Con out of Econometrics ?" American Economic Review 75, pp. 293-307.

MIRON, J.A. (1991), "Comment on Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots by Campbell and Perron," in S. Fisher ed., NBER Macroeconomics Annual 1991, Cambridge, MIT Press, pp. 141-201.

MITTNIK, S. and P.A. ZADROZNY (1993), "Asymptotic Distributions of Impulse Responses, Step Responses, and Variance Decompositions of Estimated Linear Dynamic Models," Econometrica 61, pp. 857-870.

MIZON, G.E. (1984) "The Encompassing Approach in Econometrics," in D.F. Hendry and K.F. Wallis eds., Econometrics and Quantitative economics, Basil Blackwell, pp. 135-172.

MIZON, G.E. and J.F. RICHARD (1986), "The Encompassing Principle and Its Applications to Testing Non-nested Hypotheses," Econometrica 54, pp. 657-678.

NANKERVIS, J.C. and N.E. SAVIN (1985), "Testing the Autoregressive Parameter with the t-Statistic," Journal of Econometrics 27, pp. 143-162.

NELSON, W.K. and H. KANG (1981) , "Spurious Periodicity in Inappropriately Detrended Time Series," Journal of Monetary Economics 10, pp. 139-162.

NELSON, W.K. and H. KANG (1984) , "Pitfalls in the Use of Time as an Explanatory Variable in Regression," Journal of Business and Economic Statistics 2, pp. 73-82.

NELSON, C.R. and C.I. PLOSSER (1982), "Trends and Random Walks in Macroeconomic Time Series, " Journal of Monetary Economics 10, pp. 139-162.

NEWEY, W.K. and K.D. WEST (1987), "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica 55, pp. 819-847.

NEWBOLD, P. and T. BOS (1982), "On the Use of the Hannan-Rissanen Criterion in Time Series Model Selection," mimeo., Department of Economics, University of Illinois.

NG, S. and P. PERRON (1993), "Unit Root Tests in Autoregressive-Moving Average Models with Data Dependent Methods for the Truncation Lag," mimeo., Département de Sciences économiques, Université de Montréal.

OHANIAN, L.E. (1988), "The Spurious Effects of Unit Roots on Vector Autoregressions: A Monte Carlo Study," Journal of Econometrics 39, pp. 251-266.

OSTERWALD-LENUM, M. (1992), "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," Oxford Bulletin of Economics and Statistics 54, pp. 461-472.

PAGAN, A.R. (1984) "Model Evaluation by Addition," in D.F. Hendry and K.F. Wallis eds., Econometrics and Quantitative economics, Basil Blackwell, pp. 103-134.

PAGAN, A.R. and A.D. HALL (1983), "Diagnostic Tests as Residuals Analysis", Econometrics Reviews 2, pp. 159-218.

PAGAN, A.R. and D.F. NICHOLLS (1984), "Estimating Predictions, Prediction Errors and their Standard Deviations Using Constructed Variables," Journal of Econometrics 24, pp. 293-310.

PAGAN, A.R. and M.R. WICKENS (1989) "A Survey of Some Recent Econometrics Methods," The Economic Journal 99, pp. 962-1025.

PANTULA, S.G. and A. HALL (1991), "Testing for Unit Roots in Autoregressive Moving Average Models," Journal of Econometrics 48, pp. 325-353.

PAQUET, A. (1989), "An Empirical Investigation of the Real Interest Rate Differential between Canada and the United States," CERPE, cahier de recherche no. 19.

PAQUET, A. (1992), "Inflationary Expectations and Rationality," Economics Letters 40, pp. 303-308.

PAQUET, A. (1993) Notes de cours de macroéconométrie, mimeo., 2nd edition, Université du Québec à Montréal.

PARK, J.Y. (1990), "Testing for Unit Roots and Cointegration by Variable Addition," in T.B. Fomby and G.F. Rhodes eds., Advances in Econometrics: Co-integration, Spurious Regressions, and Unit Roots, Greenwich, CT, JAI Press.

PARK, J.Y. and B. CHOI (1988) "A New Approach to Testing for a Unit Root," Cornell University, CAE WP No. 88-23.

PARK, J.Y. , S. OULIARIS and B. CHOI (1988), "Spurious Regression and Tests for Cointegration," Cornell University, CAE WP No. 88-07.

PARK, J. Y. and P.C.B. PHILLIPS (1988), "Statistical Inference in Regressions with Integrated Processes: Part I," Econometric Theory 4, pp. 468-497.

PARK, J. Y. and P.C.B. PHILLIPS (1989), "Statistical Inference in Regressions with Integrated Processes: Part II," Econometric Theory 5, pp. 95-131.

PERRON, P. (1988), "Trends and Random Walks in Macroeconomic Time Series," Journal of Economic Dynamics and Control, pp. 297-332.

PERRON, P. (1989), "The Great Crash, the Oil price Shock, and the Unit Root Hypothesis," Econometrica 57, pp. 1361-1401.

PERRON, P. (1990), "Testing for a Unit Root in a Time Series with a Changing Mean," Journal of Business and Economic Statistics 8, pp. 153-162.

PERRON, P. (1993), "Nonstationarities and Nonlinearities in Canadian Inflation," paper presented at the Bank of Canada Conference on Economic Behaviour and Policy Choice under Price Stability.

PERRON, P. and P.C.B. PHILLIPS (1987), "Does GNP Have a Unit Root? : A Re-evaluation," Economics Letters 23, pp. 139-145.

PERRON, P. and T.J. VOGELSANG (1992), "Nonstationarity and Level Shifts With an Application to Purchasing Power Parity," Journal of Business and Economic Statistics 10, pp. 301-320.

PESARAN, M.H. and B. PESARAN (1991), Microfit 3.0: User Manual, Oxford, Oxford University Press.

PESARAN, M.H., R.P. SMITH, and S. YEO (1985), "Testing for Structural Stability and Predictive Failure: a Review," Manchester School 53, pp. 280-295.

PHILLIPS, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics," Journal of Econometrics 33, pp. 311-340.

PHILLIPS, P.C.B. (1987), "Time Series Regression with a Unit Root," Econometrica 55, pp. 277-301.

PHILLIPS, P.C.B. (1991a), "Optimal Inference in Cointegrated Systems," Econometrica 59, pp. 283-306.

PHILLIPS, P.C.B. (1991b), "Unidentified Components in Reduced Rank Regression Estimation of ECM's," Yale Cowles Foundation Discussion Paper No 1003.

PHILLIPS, P.C.B. (1991c), "Special Regression for Cointegrated Time Series," in W.A. Barnett, J. Powell, and G. Tauchen eds., Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge, Cambridge University Press, pp. 413-435.

PHILLIPS, P.C.B. (1993), "Fully Modified Least Squares and Vector Autoregression," Marschak Lecture of the Far Eastern Meetings of the Econometric Society, mimeo., Cowles Foundation for Research in Economics, Yale University.

PHILLIPS, P.C.B. (1994), "Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Correction Models," Econometrica 62, pp. 73-93.

PHILLIPS, P.C.B. and S.N. DURLAUF (1986), "Multiple Time Series Regression with Integrated Processes," Review of Economic Studies 53, pp. 473-495.

PHILLIPS, P.C.B. and B.E. HANSEN (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," Review of Economic Studies 57, pp. 99-125.

PHILLIPS, P.C.B. and M. LORETAN (1991), "Estimating Long Run Economic Equilibria," Review of Economic Studies 58, pp. 407-436.

PHILLIPS, P.C.B. and S. OULIARIS (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," Econometrica 58, pp. 165-193.

PHILLIPS, P.C.B. and P. PERRON (1988), "Testing for a Unit Root in Time Series Regression," Biometrika 75, pp. 335-346.

PINDYCK, R.S. and D.L. RUBINFELD (1992), Econometric Models and Economic Forecasts, 3rd edition, McGraw-Hill.

PITARAKIS,.J.Y. (1993), "Cointegration Analysis in Large Systems: New Test Criteria and a Model Selection Procedure," mimeo., Department of Economics, Boston University.

PLOSSER, C.I. (1979), "The Analysis of Seasonal Economic Models," Journal of Econometrics 10, pp. 147-163.

PLOSSER, C.I., G.W. SCHWERT, and H. WHITE (1982), "Differencing a Test od Specification," International Economic Review 23, pp. 535-552.

PRIESTLEY, M.B. (1989), Nonlinear and Nonstationary Time Series Analysis, New York, Academic Press.

QUENOUILLE, M.H. (1949), "Approximate Tests of Correlation in Time Series," Journal of the Royal Statistical Society B11, pp. 68-94.

RACETTE, D. and J. RAYNAULD (1992), "A BVAR Model for Forecasting and Analysis of Canadian Monetary Policy," in D. Longworth ed., Monetary Seminar 90, Bank of Canada, pp. 315-385.

RAMANATHAN, R. (1989), Introductory Econometrics with Applications, Orlando, H.B.J. Publishers.

RAMSEY, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," Journal of the Royal Statistical Society Series B 31, pp. 350-371.

RAMSEY, J.B. and R. GILBERT (1972), "A Monte Carlo Study of Some Small Sample Properties of Tests for Specification Errors," Journal of the American Statistical Association 67, pp. 180-186.

REIMERS, H.E. (1993), "Comparisons of Tests for Multivariate Cointegration," Institute of Statistics and Econometrics, Working Paper No. 58, Christian-Albrechts University, Kiel.

REINSEL, G.C. and S.H. AHN (1988), "Asymptotic Distribution of the Likelihood Ratio Test for Cointegration in the Nonstationary Vector AR Model," Technical Report, University of Wisconsin, Madison.

RICE, J. (1984), "Bandwitdth Choice for Nonparametric Kernel Regression," Annals of Statistics 12, pp. 1215-1230.

RUDEBUSCH, G.D. (1990), " Trends and Random Walks in Macroeconomics Time Series: A Re-examination," Federal Reserve Paper No. 1139, Washington, D.C.

RUNKLE, D.E. (1987), "Vector Autoregressions and Reality," Journal of Business and Economic Statistics 5, pp. 437-442

SAID, E.S. and D.A. DICKEY (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika 71, pp. 599-607.

SAID, E.S. and D.A. DICKEY (1985), "Hypothesis Testing in ARIMA(p,1,q) Models," Journal of the American Statistical Association 80, pp. 369-374.

SAIKKONEN, P. (1991), "Asymptotically Efficient Estimation of Cointegrating Regressions," Econometric Theory 7, pp. 1-21.

SARGAN, J.D. (1964), "Wages and Prices in the United Kingdom: a Study in Econometric Methodology," in P.E. Hart, G. Mills and J.K. Whitaker eds., Econometric Analysis for National Economic Planning, London, Butterworths; reprinted in D.F. Hendry and K.F. Wallis eds. (1984) Quantitative Economics and Econometric Analysis, Oxford, Basil Blackwell.

SARGAN, J.D. and A. BHARGAVA (1983), "Testing the Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," Econometrica 51, pp. 153-174.

SHAPIRO, M.D. and M.W. WATSON (1988), "Sources of Business Cycle Fluctuations," in S. Fisher ed., NBER Macroeconomics Annual 1988, Cambridge, MIT Press, pp. 111-156.

SCHWARZ, G. (1978), "Estimating the Dimension of a Model," Annals of Statistics 6, pp. 461-464.

SCHWERT, G.W. (1987a), "Tests for Unit Roots: A Monte Carlo Investigation," General Working Paper Series No. 87-01, W.E. Simon Graduate School of Business Administration, University of Rochester, Rochester.

SCHWERT, G.W. (1987b), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," Journal of Monetary Economics 20, pp. 73-103.

SCHWERT, G.W. (1989), "Tests for Unit Roots : A Monte Carlo Investigation," Journal of Business and Economic Statistics 7, pp. 147-159.

SHEFFRIN, S.M. (1989), The Making of Economic Policy : History, Theory, Politics, Basil-Blackwell, ch.3.

SHIBATA, R. (1981), "An Optimal Selection of Regression Variables," Biometrika 68.

SHILLER, R.J. and P. PERRON (1985), "Testing the Random Walk Hypothesis : Power Versus Frequency of Observation," Economic Letters 18, pp. 381-386.

SIMS, C.A. (1980), "Macroeconomic and Reality," Econometrica 48, pp. 1-47.

SIMS, C.A. (1986), "Are Forecasting Models Usable for Policy Analysis?," Federal Reserve Bank of Minneapolis Quarterly Review 10, pp. 2-16.

SIMS, C.A. (1988), "Bayesian Skepticism on Unit Root Econometrics," Journal of Economic Dynamics and Control 12, pp. 463-474.

SIMS, C.A., J. STOCK and M.W. WATSON (1990), "Inference in Linear Time Series Models with Some Unit Roots," Econometrica 58, pp. 113-144.

SPENCER, D. and K. BERK (1981), "A Limited Information Specification Test," Econometrica 49, pp. 1079-1085.

STIGLER, S.M. (1981), "Gauss and the Invention of Least Squares," The Annals of Statistics 9, pp. 465-474.

STOCK, J.H. (1987), "Asymptotic properties of Least Squares Estimators of Cointegrating Vectors," Econometrica 55, pp. 1045-1056.

STOCK, J.H. and M.W. WATSON (1988), "Variable Trends in Economic Time Series," Journal of Economic Perspectives 2, pp. 147-174.

STOCK, J.H. and M.W. WATSON (1989), "A Simple MLE of Cointegrating Vectors in Higher Order Integrated Systems," National Bureau of Economic Research Technical Working Paper No. 83, Cambridge.

THEIL, H. (1961), Economic Forecasts and Policy, 2nd edition, Amsterdam, North-Holland.

TODA, H.Y. and P.C.B. PHILLIPS (1993a), "The Spurious Effect of Unit Roots on Vector Autoregressions," Journal of Econometrics 59, pp.229-255.

TODA, H.Y. and P.C.B. PHILLIPS (1993b), "Vector Autoregressions and Causality," Econometrica 61, pp. 1367-1393.

TODA, H.Y. and P.C.B. PHILLIPS (1994), "Vector Autoregression and Causality: A Theoretical Overview and Simulation Study," Econometric Reviews 13, pp. 259-285.

WATSON, M.W. (1993), "Vector Autoregressions and Cointegration", Working Papers Series No. 93-14, Macroeconomic Issues, Research Department, Federal Reserve Bank of Chicago, forthcoming in R.F. Engle and D. McFadden eds., Handbook of Econometrics, Vol. 4.

WEST, K.D. (1988a), "Asymptotic Normality when Regressors have a Unit Root," Econometrica 56, pp. 1394-1417.

WEST, K.D. (1988b), "On the Interpretation of Near Random Walk Behavior in GNP." American Economic Review 78, pp. 202-209.

WHITE, H. (1980), "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," Econometrica 48, pp. 817-838.

WHITE, H. (1984), Asymptotic Theory for Econometricians, Orlando, Academic Press.

WICKENS, M.W. and T.S. BREUSCH (1988), "Dynamic Specification, the Long-Run and the Estimation of Transformed Regression Models," Economic Journal 98, pp. 189-205.

WICHERN, D.W. (1973), "The Behavior of the Sample Autocorrelation Function for an Integrated Moving Average Process," Biometrika 60, pp. 235-239.

WOOLRIDGE, J.M. (1991), "Notes on Regression with Difference-Stationary Data," mimeo., Michigan State University.

ZELLNER, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association 57, pp. 348-368.

ZIVOT, E. and D.W.K. ANDREWS (1992), "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," Journal of Business and Economic Statistics 10, pp. 251-270.

LIST OF WORKING PAPERS

94-5 A Guide to Applied Modern Macroeconometrics by **Alain Paquet**

94-4 Interindustry Wage Differentials: Evidence, Explanations and Potential Distortions by **Mario Iacobacci**

94-3 Regional Productivity Convergence in Canada by **Frank C. Lee and Serge Coulombe**

94-2 Regional Economic Disparities in Canada by **Frank C. Lee and Serge Coulombe**

94-1 The Interaction of Inflation with a Canadian-Type Capital Tax System: a Dynamic General Equilibrium Analysis Incorporating Endogenous Leverage and Heterogeneous Households by **Steven James**

92-4 An Environmental CGE Model of Canada and the United States by **Louis Beauséjour, Gordon Lenjosek and Michael Smart**

92-3 Estimating Trend Total Factor Productivity Growth in Canada by **Carolyn Wilkins, Frank Lee and Steven James**

92-2 Un indice de prix régional de biens et services comparables au Canada et son application aux disparités régionales de revenue par **Raynald Létourneau**

92-1 A New Composite Leading Indicator of the Canadian Economy by **Robert Lamy**

90-7 Une analyse empirique des déterminants du taux d'intérêt réel mondial par **Benoît Robidoux**

90-6 Real Interest Rates in Historical Perspective by **Paul Jenkins**

90-5 The Contribution of Fiscal Policy to Real Interest Rate Movements in the 1980s by **Paul Jenkins**

90-4 The Constant Price Hypothesis in Multi-Industry Cross-Section Studies: The Impact on Cost Function Estimates by **Benoît Robidoux and John Lester**

90-3 Regional Disparities in Canadian Manufacturing: The Impact of Plant Size and Technology by **Benoît Robidoux and John Lester**

90-2 The Determination of Real Interest Rates by **Joel Fried and David Burgess**

90-1 La voie des dépenses fiscales: le point sur les questions conceptuelles et les controverses par **Neil Bruce et Michael Daly**

90-1 Pathways to Tax Expenditures: A Survey of Conceptual Issues and Controversies by **Neil Bruce**

89-7 INTERMOD 2.0: Model Specification and Simulation Properties by **Guy Meredith**

89-6 The Calculation of Marginal Effective Tax Rates in the 1987 White Paper on Tax Reform by **Jack Jung**

89-5 Effects of Changing Age Structure on Consumption and Savings by **Paul Jenkins**

89-4 Expectations, Policy Shocks, and Output Behaviour in a Multi-Country Macroeconomic Model by **Guy Meredith**

89-3 The Goods and Services Tax: A General Equilibrium Analysis by **Bob Hamilton and Chun-Yan Kuo**

89-2 A Survey of Some Recent Developments in Macroeconomics by **Michael Devereux**

89-1 Closed Economy Analysis of the Dynamics of Deficits and Debt in a Unitary and Federal State by **Carolyn Wilkins and Mario Albert**

88-10 A Guide to the Get Model by **Richard G. Harris**

88-9 Intermod Model User's and Developer's Guide for PC Systems by **P. Bagnoli and A. Kruns**

88-8 The Trade Sector in INTERMOD by **Guy Meredith**

88-7 Intermod 1.1: A G-7 Version of the IMF's Multimod by **John Helliwell, Guy Meredith, P. Bagnoli and Yves Durand**

88-6 Resolutions of the U.S. Trade Imbalances: How Painful Will the Adjust Be? by **Janice Haritos**

88-5 Econometric Estimates of Import Price Elasticities for Canada by **Raynald Létourneau and John Lester**

88-4 Econometric Estimates of Scale Economies in Canadian Manufacturing by **Benoît Robidoux and John Lester**

88-3 Trade Barriers Between Canada and the United States by **John Lester and Tony Morehen**

88-2 New Estimates of Canadian Tariff Rates by Industry and Commodity by **John Lester and Tony Morehen**

88-1 Geographic Labour Mobility in Canada by **Louis Grignon and Kei Moray**

87-1 Économies d'échelle: Un survol des techniques d'estimation et des problèmes méthodologiques par **Benoît Robidoux**

86-1 The Structure of Production in Ten Canadian Industries by **François Delorme and John Lester**

2001	
2001-05	Monitoring Regional Economies in Canada With New High-Frequency Coincident Indexes - Robert Lamy and Patrick Sabourin
2001-04	Labour or Total Factor Productivity: Do We Need to Choose? - Timothy C. Sargent and Edgard R. Rodriguez
2001-03	Structural Unemployment and Technical Change in Canada, 1990-1999 - Timothy C. Sargent
2001-02	Determinants of Retirement: Does Money Really matters? - Janice Compton
2001-01	The role of Government Debt in a World of Incomplete Financial Markets - Steven James and Philippe Karam
2000	
2000-08	Public Finance Implications of Population Ageing - Phillip King and Harriet Jackson
2000-07	Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models – Steven Gonzalez
2000-06	Forecasting employment Rates: A Cohort Approach – Marie-France Paquet and Timothy C. Sargent
2000-05	Measuring Human Capital in Canada - Mireille Laroche and Marcel Mérette of the University of Ottawa
2000-04	L'écart à terme comme indicateur de l'activité économique : L'expérience des années 90 Paul Delâge et Jean-François Fillion
2000-03	Croissance endogène et vieillissement démographique : le cas d'une petite économie ouverte* Maxime Fougère
2000-02	"Modeling Federal Finances under Uncertainty" Derek Hermanutz and Chris Matier
2000-01	"Liability Management Using Dynamic Portfolio Strategies" Richard Black, Colin R. Telmer
1999	

99-03 "An Econometric Examination of the Impact of Population Ageing on Personal Savings in Canada" , Maxime Fougère - Finance, Marcel Mérette - University of Ottawa

99-02 "Analyzing and Forecasting Credit Ratings: Some Canadian Evidence" Patrick Sabourin

99-01 "Taxation and Economic Performance: A Cross-Country Comparison and Model Sensitivity Analysis" Jing Xu

1998

98-09 "The BU Ratio: Prospect and Retrospect" Timothy C. Sargent

98-08 "Youth and the 1990s Labour Market" Carolyn Wilkins

98-07 "The Canadian Economic and Fiscal Model - 1996 Version: Part 3 - Empirical Specification and Statistical Assessment" Jian-Guo Cao and Benoît Robidoux

98-06 "The Canadian Economic and Fiscal Model - 1996 Version: Part 2 - Dynamic Forecasting and Simulation Properties" Alain DeSerres, Benoît Robidoux and Bing-Sun Wong

98-05 "The Canadian Economic and Fiscal Model - 1996 Version: Part 1 - Model Structure" Benoît Robidoux and Bing-Sun Wong

98-04 "Population Ageing and the Current Account in Selected OECD Countries" Maxime Fougère and Marcel Mérette

98-03 "Population Ageing and Economic Growth in Seven OECD Countries" Maxime Fougère and Marcel Mérette

98-02 "The Persistence of Low Income Spells in Canada, 1982-1993" Mireille Laroche

98-01 "On the Concept and Dimensions of Human Capital in a Knowledge-Based Economy Context" Mireille Laroche, Marcel Mérette and G.C. Ruggeri

1997

97-08 "Estimating the Economic Efficiency Consequences of Alternative Tax Reforms Using a CGE Model: Homogeneous Vs Heterogeneous Labour Specifications" Louis Beauséjour, Mokhtar Souissi, Carole Vincent and G.C. Ruggeri

97-07 "Tax Expenditures and Tax Preferences in the Personal Income Tax System of Selected OECD Countries: A Suggested Classification, G.C. Ruggeri and Carole Vincent

97-06 "Income Taxes, Life-Cycles and Growth", Marcel Mérette

97-05 "Effective Income Tax Rates in Macroeconomics: An Update for Canada", G.C. Ruggeri, Mireille Laroche and Carole Vincent

97-04 "A Public versus a Private Canada Pension Plan: A Survey of the Economics" Steven James

97-03 "Forecasting Canadian Recessions with Macroeconomic Indicators" Robert Lamy

97-02 "The Dynamic Effects of Taxes and Government Spending in a Calibrated Canadian Endogenous Growth Model" Jing Xu

97-01 "The Impact of Fiscal Policy on the Risk Premium of Government Long-Term Debt: Some Canadian Evidence" Todd Mattina and François Delorme

1996

96-06 "Does the Length of the Current Expansion Increase the Probability of Recession?" Lisa Backa

96-02 "Do Mechanical Filters Provide a Good Approximation of Business Cycles?", Alain Guay and Pierre St-Amant

96-01 "An Indicator Model of Core Inflation in Canada" Robert Lamy and Paul Rochon

1995

95-12 "Potential Economic Effects of Experience-Rating the Unemployment Insurance System Using a Multi-Sector General Equilibrium Model of Canada" Louis Beauséjour, Munir A. Sheikh and Baxter Williams

95-11 "Active Labour Market Policies: An Evaluation" Lori Marchildon

95-10 "An Index of Unemployment Insurance Disincentives" Timothy C. Sargent

95-09 "The Economics of Canada Pension Plan Reforms" Steven James, Chris Matier, Humam Sakhnini and Munir Sheikh

95-08 "The Long-Run Economic Impacts of Government Debt Reduction" Steven James and Chris Matier

95-07 "Time Series Properties of Canadian Real Interest Rates", Chris Matier and Werner Schüle

95-06 "Interpreting Sacrifice Ratios Across Countries and Over Time" Brenda Lipsett and Steven James

95-05 "Effective Rates of Assistance: An Old Idea Whose Time Has Come?" Louis Beauséjour and Munir A. Sheikh

95-04 "Aggregate Fluctuations, Monetary Policy and Nominal Wage Rigidities in a New-Classical Setting: Results for Canada" Humam Sakhnini

95-03 "Ricardian Equivalence and the Permanent Income Hypothesis: An Empirical Investigation" Benoit Perron

95-02 "Are New-Classical Models Becoming Relevant Policy Tools: A Survey of Recent New-Classical Models" Humam Sakhnini

95-01 "Fluctuations économiques dans un environnement de stabilité des prix" Claude Lavoie

Please note that this working paper is available in French only.

1994

94-06 "Conditional Productivity Convergence in Canada" Frank C. Lee

94-05 "A Guide to Applied Modern Macroeconometrics" Alain Paquet

94-04 "Interindustry Wage Differentials: Evidence, Explanations and Potential Distortions" Mario Iacobacci

94-03 "Regional Productivity Convergence in Canada" Frank C. Lee and Serge Coulombe

94-02 "Regional Economic Disparities in Canada" Serge Coulombe and Frank C. Lee

94-01 "The Interaction of Inflation with a Canadian-Type Capital Tax System: A Dynamic General Equilibrium Analysis Incorporating Endogenous Leverage and Heterogeneous Households" Steven James

1992

92-04 "An Environmental CGE Model of Canada and the United States" Louis Beauséjour, Gordon Lenjosek and Michael Smart

92-03 "Estimating Trend Total Factor Productivity Growth in Canada" Carolyn Wilkins, Frank Lee and Steven James

92-02 "Un indice de prix régional de biens et services comparables au Canada et son application aux disparités régionales de revenu" Raynald Létourneau. Available in French only.

92-01 "A New Composite Leading Indicator of the Canadian Economy" Robert Lamy

1990

90-07 "Une analyse empirique des déterminants du taux d'intérêt réel mondial" Benoît Robidoux
Please note that this working paper is available in French only.

90-06 "Real Interest Rates in Historical Perspective" Paul Jenkins

90-05 "The Contribution of Fiscal Policy to Real Interest Rate Movements in the 1980s" Paul Jenkins

90-04 "The Constant Price Hypothesis in Multi-Industry Cross-Section Studies: The Impact on Cost Function Estimates" Benoît Robidoux and John Lester

90-03 "The Cyclically-Adjusted Output Methodology" Mario Albert, Nathalie Boisvert and Robert Fonberg

90-02 "The Determination of Real Interest Rates" Joel Fried and David Burgess

90-01 "La voie des dépenses fiscales : le point sur les questions conceptuelles et les controverses"
Neil Bruce. Available in French only.

1989

89-07 "INTERMOD 2.0: Model Specification and Simulation Properties" Guy Meredith

89-06 "The Calculation of Marginal Effective Tax Rates in the 1987 White Paper on Tax Reform"
Jack Jung

89-05 "Effects of Changing Age Structure on Consumption and Saving" Paul Jenkins

89-04 "Expectations, Policy Shocks, and Output Behaviour in a Multi-Country Macroeconomic Model"
Guy Meredith

89-03 "The Goods and Services Tax: A General Equilibrium Analysis" Bob Hamilton and Chun-Yan Kuo

89-02 "A Survey of Some Recent Developments in Macroeconomics" Michael Devereux

89-01 "Closed Economy Analysis of the Dynamics of Deficits and Debt in a Unitary and Federal State" - Carolyn Wilkins and Mario Albert

1988

88-10 "A Guide to the GET Model" Richard G. Harris

88-09 "INTERMOD Model User's and Developer's Guide for PC Systems" Philip Bagnoli and Anders Kruus

88-08 "The Trade Sector in INTERMOD" Guy Meredith

88-07 "INTERMOD 1.1: A G-7 Version of the IMF's Multimod" John F. Helliwell, Guy Meredith, Yves Durand and Philip Bagnoli

88-06 "Resolutions of the U.S. Trade Imbalances: How Painful Will the Adjust Be?" Janice Haritos

88-05 "Econometric Estimates of Import Price Elasticities for Canada" Raynald Létourneau and John Lester

88-04 "Econometric Estimates of Scale Economies in Canadian Manufacturing" Benoît Robidoux and John Lester

88-03 "Trade Barriers Between Canada and the United States" John Lester and Tony Morehen

88-02 "New Estimates of Canadian Tariff Rates by Industry and Commodity" John Lester and Tony Morehen

88-01 "Geographic Labour Mobility in Canada" Louis Grignon and Kei Moray

1987

87-01 "Économies d'échelle : Un survol des techniques d'estimation et des problèmes méthodologiques"
Benoît Robidoux. Available in French only.

1986

86-01 "The Structure of Production in Ten Canadian Industries" François Delorme and John Lester